

**НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ**  
**«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ**  
**імені ІГОРЯ СІКОРСЬКОГО»**

**Інститут прикладного системного аналізу**  
**Кафедра математичних методів системного аналізу**

«На правах рукопису»  
УДК 519.854

«До захисту допущено»  
Завідувач кафедри  
\_\_\_\_\_ О.Л. Тимощук  
«\_\_» \_\_\_\_\_ 20\_\_ р.

**Магістерська дисертація**

**на здобуття ступеня магістра**  
**за освітньо-науковою програмою**  
**зі спеціальності 124 Системний аналіз**

**на тему: «Система підтримки прийняття рішень для програмного комплексу**  
**автоматизованого формування розкладу сеансів кінотеатрів»**

Виконала:

студентка II курсу, групи КА-92мп

Зінченко Анастасія Олександрівна \_\_\_\_\_

Керівник:

професор кафедри ММСА, д.т.н.

Мухін Вадим Євгенович \_\_\_\_\_

Рецензент:

доцент кафедри технічної кібернетики

КПІ ім. Ігоря Сікорського, д.т.н., доцент

Корнага Я.І. \_\_\_\_\_

Засвідчую, що у цій  
магістерській дисертації немає  
запозичень з праць інших авторів без  
відповідних посилань.

Студентка \_\_\_\_\_

**КИЇВ**

**2020**

**Національний технічний університет України**  
**«Київський політехнічний інститут**  
**імені Ігоря Сікорського»**  
**Інститут прикладного системного аналізу**  
**Кафедра математичних методів системного аналізу**

Рівень вищої освіти – другий (магістерський)

Спеціальність – 124 «Системний аналіз»

ЗАТВЕРДЖУЮ

Завідувач кафедри

\_\_\_\_\_

(підпис)

\_\_\_\_\_

(ініціали, прізвище)

«\_\_» \_\_\_\_\_ 20\_\_ р.

### ЗАВДАННЯ

**на магістерську дисертацію студентці**

**Зінченко Анастасії Олександрівни**

1. Тема дисертації «Система підтримки прийняття рішень для програмного комплексу автоматизованого формування розкладу сеансів кінотеатрів», науковий керівник дисертації Мухін Вадим Євгенович, д.т.н., професор, затверджені наказом по університету від «02» листопада 2020 р. № 3182-с. Термін подання студентом 12.12.20
3. Об'єкт дослідження: Задача автоматизації основних виробничих процесів кінотеатрів.
4. Предмет дослідження: Метод автоматичного формування розкладу сеансів кінотеатрів.
5. Перелік завдань, які потрібно розробити:
  1. Здійснити огляд технічної літератури за темою роботи;
  2. Дослідити актуальність обраної теми;
  3. Ознайомитись із існуючими методами автоматичного формування розкладу сеансів кінотеатрів;
  4. Здійснити порівняльний аналіз наявних методів, проаналізувати їх переваги та недоліки;
  5. Запропонувати і реалізувати метод автоматичного формування розкладу сеансів кінотеатрів;
  6. Провести аналіз результатів;
  7. Провести аналіз ринкових можливостей запуску стартап проекту;

8. Розробити концептуальні висновки;
  9. Підготувати ілюстративний матеріал;
  10. Оформити пояснювальну записку.
6. Орієнтовний перелік ілюстративного матеріалу:
1. Постановка завдання дослідження;
  2. Існуючі методи автоматичного формування розкладу сеансів кінотеатрів;
  3. Структура програмного продукту;
  4. Таблиці з результатами моделей.
7. Орієнтовний перелік публікацій:
1. Сидоров С. Г., Никологорская А. В. Анализ временных рядов как метод построения потребления электроэнергии / Вестник ИГЭУ. 2010, Вып. 3. С. 81–83.
  2. Бокс Дж., Дженкинс Г. М. Анализ временных рядов, прогноз и управление. Москва: Мир, 1974. 406 с.
  3. Фейгин Л. И. Задачи теории расписаний при нечетких длительностях операций. Доклады АН СССР. 1983. Т. 272. №4. С. 812–815.
8. Дата видачі завдання: 5 вересня 2019 року.

## Календарний план

№ з/п	Назва етапів виконання магістерської дисертації	Термін виконання етапів магістерської дисертації	Примітка
1.	Отримання завдання на магістерську дисертацію	01.09.2020 – 06.09.2020	
2.	Огляд технічної літератури за темою	07.09.2020 – 13.09.2020	
3.	Концептуальний вступ дисертації. Формулювання об'єкта, предмета, цілі, завдань, новизни, практичної значущості результатів. Дослідження актуальності обраної теми	14.09.2020 - 20.09.2020	
4.	Перший розділ. Аналіз існуючих методів для автоматичного формування розкладу сеансів кінотеатрів	21.09.2020 - 27.09.2020	
5.	Другий розділ. Розробка нового методу для автоматичного формування розкладу сеансів кінотеатрів	28.09.2020 - 04.10.2020	
6.	Розробка початкового програмного забезпечення	05.10.2020 – 11.10.2020	
7.	Третій розділ. Опис структури побудованого програмного комплексу для автоматичного формування розкладу сеансів кінотеатрів	12.10.2020 - 18.10.2020	
8.	Аналіз результатів	19.10.2020 – 25.10.2020	
9.	Проведення аналізу ринкових можливостей стартап-проекту	26.10.2020 - 01.11.2020	
10.	Підготовка ілюстративного матеріалу	02.11.2020 - 15.11.2020	
11.	Оформлення пояснювальної записки	16.11.2020 – 27.11.2020	

Студент \_\_\_\_\_  
 Науковий керівник дисертації \_\_\_\_\_

А.О. Зінченко  
 В.Є.Мухін

## РЕФЕРАТ

Магістерська дисертація містить 83 с., 28 табл., 30 рис., список використаних джерел із 16 найменувань, 1 дод.

КІНОТЕАТР, МОДЕЛЮВАННЯ, НЕЙРОННА МЕРЕЖА, ПРОГНОЗУВАННЯ, ПАРСИНГ ДАНИХ, РОЗКЛАД.

Об'єкт дослідження – Задача автоматизації основних виробничих процесів кінотеатрів.

Предмет дослідження - Метод автоматичного формування розкладу сеансів кінотеатрів.

Мета роботи - Покращити програмний комплекс автоматизованого формування розкладу сеансів для кінотеатрів на основі статистичних даних та даних схожих фільмів.

Методи дослідження – системного математичного аналізу, математичного прогнозування, математичного моделювання, статистичні.

В ході виконання роботи розроблений метод автоматизованого формування розкладу сеансів кінотеатрів на основі статистичних даних, що складається з п'яти етапів:

1. Очистка та нормалізація вхідних даних;
2. Аналіз історії прокату фільму (якщо є);
3. Прогнозування кількості людей на можливий сет сеансів;
4. Аналіз додаткової інформації про сеанс;
5. Прогнозування можливого розкладу з можливого сету сеансів на 1 день.

Під час аналізу предметної області були виявлені її слабкі місця, а дослідження аналогів показало, що жодна з програм не має в собі потрібного функціоналу.

Для створення цього комплексу було проаналізовано основні різновиди математичних моделей та методів, розглянуто принципи побудови ансамблів, беггінгу та бустингу, досліджено приклади використання мереж з довгостроковою пам'яттю для вирішення схожого типу задач а також проведено детальний аналіз області кінематографії.

В результаті роботи створено програму що відповідає поставленим вимогам, а через її модульність всі її частини можуть бути використані окремо для інших проектів в цій та схожих областях.

Ця програма має допомогти власникам кінотеатрів спростити процес формування розкладу та забезпечити підвищення прибутку за рахунок аналізу статистичних даних схожих фільмів, що були в прокаті раніше та мета-даних нових фільмів.

Сфера застосування – кінопрокат.

## ABSTRACT

The master's dissertation contains 83 pages, 28 tables, 30 figures, the list of the used sources from 16 names, 1 appendix.

CINEMA, SIMULATION, NEURAL NETWORK, FORECASTING, DATA PARSING, SCHEDULE.

Object of research - The task of automation of the main production processes of cinemas.

Subject of research - Method of automatic formation of the schedule of sessions of cinemas.

Purpose - To improve the software package of automated scheduling for cinemas based on statistics and data from similar films.

Research methods - system mathematical analysis, mathematical forecasting, mathematical modeling, statistical.

During the work the method of automated formation of the schedule of cinema sessions on the basis of statistical data consisting of five stages is developed:

1. Cleaning and normalization of input data;
2. Analysis of the film's rental history (if any);
3. Forecasting the number of people for a possible set of sessions;
4. Analysis of additional information about the session;
5. Forecasting a possible schedule from a possible set of sessions for 1 day.

The analysis of the subject area revealed its weaknesses, and the study of analogues showed that none of the programs has the necessary functionality.

To create this complex, the main types of mathematical models and methods were analyzed, the principles of ensemble building, begging and boosting were considered, examples of using networks with long-term memory to solve similar problems were studied, and a detailed analysis of cinematography was performed.

As a result of work the program meeting the set requirements is created, and because of its modularity all its parts can be used separately for other projects in this and similar areas.

This program should help cinema owners simplify the scheduling process and increase profits by analyzing statistics from similar films that have been rented before and meta-data from new films.

Scope - film distribution.

## ЗМІСТ

ВСТУП.....	8
РОЗДІЛ 1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ.....	9
1.1 Постановка задачі.....	9
1.2 Програмування модулів системи .....	9
1.3 Огляд існуючих програмних продуктів .....	10
1.4 Особливості поставленої задачі .....	12
1.5 Моделі та методи прогнозування .....	13
1.6 Модель Random forest.....	15
1.7 Градієнтний бустинг.....	17
1.8 Моделі глибокого машинного навчання .....	19
1.9 Матеріали для навчання .....	20
1.10 Рекурентні нейронні мережі .....	21
1.11 Моделі з довгою короткостроковою пам'яттю .....	22
1.12 Вимоги до програмного забезпечення .....	25
Висновки до розділу 1 .....	25
Розділ 2 РОЗРОБКА НОВОГО МЕТОДУ .....	27
2.1 Первинний аналіз вхідних даних .....	27
2.2 Обґрунтування нового методу на формально-теоретичному рівні .....	27
2.3. Теоретичні оцінки параметрів розробленого методу.....	30
Висновки до розділу 2 .....	39
Розділ 3 ОПИС РОЗРОБЛЕНОГО ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ .....	40
3.1 Структура програмного комплексу .....	40
3.2 Опис розробки програмного комплексу .....	41
Висновки до розділу 3 .....	52
Розділ 4 ОПИС ВИПРОБУВАНЬ ПРОДУКТУ .....	53
4.1 Сервіс для підбору схожих фільмів.....	53
4.2 Програмний комплекс для формування розкладу .....	55
Висновки до розділу 4.....	60
РОЗДІЛ 5. РОЗРОБКА СТАРТАП-ПРОЕКТУ .....	61
Вступ .....	61
5.2 Опис ідеї стартапу .....	62
5.3 Проектування стартапу та формування команди.....	65
5.4 Розроблення MVP .....	68
5.5 Аналіз ринкових можливостей та розроблення маркетингової стратегії стартап-проекту.....	69
5.6 Формування маркетингової програми та плану маркетингу стартапу .....	72
Висновки до розділу 5 .....	76
ВИСНОВКИ .....	78
ПЕРЕЛІК ПОСИЛАНЬ.....	79
ДОДАТОК А.....	81

## ВСТУП

У наш час важливою є тема оптимізації виробничого процесу, економії часу на планування та розподіл робіт, тощо.

Маючи доступ до бази даних підприємства можна врахувати багато факторів, провести візуальний аналіз даних, побудувати базові моделі користувацької поведінки за допомогою РФМ аналізу, тощо.

Але такий аналіз є достатньо поверхневим, він також приносить прибуток, але рано чи пізно підприємець починає шукати більш глибокі методи аналізу даних та використовувати їх. Зазвичай під час такого аналізу з'являються додаткові фактори впливу, які до цього не здавалися важливими. Це призводить до значного розширення даних аналізу, і обійтись без програмного забезпечення вже не можливо.

Область аналізу даних та прогнозування за допомогою машинного навчання в Україні має стимул до розвитку через те, що або необхідне програмне забезпечення відсутнє на українському ринку, або ж воно коштує занадто дорого та не вирішує поставлених проблем в повній мірі, а створює нові через складність у налаштуванні та використанні.

До того ж, враховуючи ситуацію у країні пов'язану з коронавірусними обмеженнями та, як наслідок, швидкою діджиталізацією компаній, проблема детального, глибокого аналізу даних є актуальною, бо дозволить швидше підняти бізнес на більш менш стабільний рівень.

Наукова новизна : Запропоновано метод формування розкладу сеансів кінотеатрів, який відрізняється від існуючих тим, що в ньому прогнозування популярності фільму в прокаті виконується на основі апарату нейронних мереж, що дозволяє виконувати коригування розкладу лише за потреби змін та без його повторного створення.



## РОЗДІЛ 1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ

Мета роботи: Покращити програмний комплекс автоматизованого формування розкладу сеансів для кінотеатрів на основі статистичних даних та даних схожих фільмів.

### 1.1 Постановка задачі

Для досягнення поставленої мети були вирішені наступні завдання:

- дослідити створену раніше систему формування розкладу сеансів для кінотеатрів, виявити слабкі місця системи розподілу сеансів;
- робити нову модель для автоматизованого формування розкладу сеансів для кінотеатрів;
- розробити програмний комплекс автоматизованого формування розкладу сеансів для кінотеатрів, враховуючи дані фільмів;
- протестувати програмний комплекс автоматизованого формування розкладу сеансів для кінотеатрів на реальних даних.

### 1.2 Програмування модулів системи

Модульність — основний принцип побудови комп'ютерних систем. Програмний модуль — це окрема функціонально закінчена програмна одиниця, яка так чи інакше ідентифікується і використовується в програмі. Модульний аналіз дозволяє виконувати роздільне програмування окремих частин програми проекту і полегшує тестування проекту в цілому [1], [2], [3].

Процес програмування заснований на ітеративному принципі: спочатку кодуються найбільш важливі елементи алгоритму, а потім відбувається додавання деталей і сервісних можливостей.

Програмні модулі повинні відповідати мінімальним вимогам:

- можливість розширення;
- адаптивність;
- портативність.

Під розширюваністю слід розуміти той факт, що при створенні нових доповнень немає необхідності перезаписувати старі.

Адаптивність програмного модуля повинна гарантувати продуктивність програмного забезпечення в цілому [4], [5].

У свою чергу, переносимість – це здатність модуля функціонувати в програмному та апаратному середовищі з варіаціями своїх властивостей. Іншими словами, модуль повинен бути незалежним від комп'ютера і конкретної операційної системи [1], [3].

### 1.3 Огляд існуючих програмних продуктів

Серед існуючих програмних реалізацій, які хоча б частково вирішують проблему є такі відомі програми як «Прем'єра» від R-keeper і KinoPlan.

Ці програми призначені для спрощення розробки плану сеансу, але частка праці оператора все ще дуже велика – програми можуть коригувати заповнений вручну розклад, перевіряючи обмеження на зали і технології, але не вміють формувати хоча б базову версію розкладу.

Програма «Прем'єра» – це готовий сервіс, в якому реалізовані модулі для бронювання квитків, відстеження завантаженості залів, інтеграції з SMS-розсилками, контролю прибутку і т. д.

Серед переваг даної послуги можна виділити наступні:

- готова до розширення система з модульною структурою;
- при створенні розкладу оператором система перевіряє існуючі обмеження;
- внутрішня база даних прокату фільмів, в якій відображаються дані про стрічку.

Серед основних мінусів – розклад вводить Оператор, а це займає багато часу. Розклад не є оптимальним, деякі зали пустують, тощо.

Програма «Кіноплан» – більш цікавий інструмент для планування. На додаток до функцій, перелічених у Прем'єрі, Kinoplan має можливість автоматичного планування відповідно до обмежень, безліч опцій для настройки показів і багато іншого.

Переваги цієї послуги більш значні:

- можливість попереднього формування розкладу;
- велика кількість налаштувань;
- автоматичне оновлення інформації про кількість обмежень від дистриб'юторів фільмів;
- візуально приємний інтерфейс;
- багато інформації про використання програми;
- модульність системи.

Однак планування без обмежень як і раніше вимагає великої участі оператора.

Оцінка самого фільму, пошук попередніх оцінок, спроби подумати за глядача і вгадати, де саме додавати сеанси, щоб отримати більший прибуток – на такі запити немає відповіді ні в одному з аналогів.

Тому метою роботи було покращити програмний комплекс автоматизованого формування розкладу сеансів для кінотеатрів на основі статистичних даних та даних схожих фільмів.

## 1.4 Особливості поставленої задачі

Принцип розподілу сеансів зараз доволі простий – вгадати на який час і в який зал прийде більше людей на конкретний фільм, та поставити туди його сеанс. Це доволі просто зробити коли фільмів близько 5 – 6, але в реальному оточенні це перетворюється на затратну по часу та ресурсам проблему.

Для того щоб мати уявлення про кількість глядачів на сеансі – потрібно знати хоча б частково інформацію про фільм. Ідеально було б знати скільки людей прийшло на нього вчора або позавчора, які були враження від перегляду, в який час було більше людей, тощо. Але є декілька нюансів. По-перше, такі дані відсутні для прем'єрних фільмів, та й кількість глядачів по вихідним та робочим дням буде сильно відрізнятися.

А по-друге, які об'єми інформації потрібно кожного разу аналізувати людині-оператору, та скільки робочого часу на це витрачається не раціонально.

Відсутність єдиної бази фільмів також сильно підводить – у різних кінотеатрах фільм може називатися по різному, може відрізнятися прокатна інформація, дозволений вік для перегляду, тощо.

Також що до вікових обмежень (табл. 1.1), то вони значно відрізняються від міжнародного стандарту, їх більше та вони часто пересікаються між собою, що не дає змоги однозначно класифікувати фільм.

Таблиця 1.1 – порівняння систем обмежень

Американська система	Наша система
G	0+, 1+, 2+, 3+
PG	6+, 7+
PG13	12+, 14+
R	18+, 21+

Крім того, ще даний перелік необхідних умов від кінотеатру, а саме :

- встановлений час роботи залів;
- різноманіття технологій в залах, обмеження на них;
- часові рамки від дистриб'юторів – ранок, день, вечір.

До того ж, є фільми які йдуть повтором, або являються ремейками – теоретично на них можна дивитись минулу історію, якщо знати фільм оригінал.

Наразі всі ці обмеження перевіряються вручну.

### 1.5 Моделі та методи прогнозування

Система прогнозування – упорядкований набір методик, технічних засобів, призначений для прогнозування складних явищ і процесів [6], [7].

Математична модель – це або фізичне уявлення математичних понять, або математичне уявлення реальності [8].

Математична модель як фізичне уявлення включає в себе відтворення твердих тіл з картону, дерева, пластика або інших матеріалів і не є предметом нашого дослідження.

Набагато ширше концепція математичної моделі як способу представлення реальності за допомогою формул і математичних понять.

Загалом, все в фізичному або біологічному світі, незалежно від того, чи є воно природним або створено під впливом технологій і людини, можна проаналізувати за допомогою математичних моделей, якщо це можна описати в математичних термінах [1], [9].

Моделі предметної області – це математичні моделі прогнозування, для побудови яких використовуються відомі закони обраної предметної області. Для розробки таких моделей необхідний індивідуальний підхід і фахівець в обраній області [2].

Моделі часових рядів – це моделі, які можна використовувати для визначення залежності майбутніх значень від минулих значень в середині процесу і через цю залежність для розрахунку прогнозу. Такі моделі універсальні, їх зовнішній вигляд не змінюється через характер часових рядів [2].

Що стосується класифікації моделей (рис. 1.1), то в моделях предметної області таке уявлення неможливо – стільки областей, скільки моделей. Навпаки, моделі часових рядів можна розділити на статистичні та структурні моделі [2], [10].



Рисунок 1.1 – Класифікація методів і моделей прогнозування

У статистичних моделях залежність минулого від майбутнього задається у вигляді рівняння [2]. До них відносяться такі моделі як:

- регресійні моделі;
- авторегресійні моделі;
- моделі експоненційного згладжування (модель Холта–Вінтерса, модель Брауна).

У структурних моделях залежність майбутньої цінності від минулого задається у вигляді певної структури і правил переходу до неї [2], [11], [12].

Приклад:

- моделі на основі нейронних мереж;
- моделі на основі ланцюгів Маркова;
- моделі на основі дерев класифікації і регресії.

В роботі використано наступні моделі:

- модель на базі класифікаційно–регресійних дерев (Random forest);
- модель на базі нейронних мереж (LSTM).

## 1.6 Модель Random forest

Модель Random forest є ансамблем множини дерев рішень, що дозволяє знизити проблему перенавчання і підняти точність в порівнянні з одним деревом. Прогнозом є результат агрегування відповідей всіх дерев.

Тренування дерев проходить на різних підмножинах, що вирішує проблему побудови однакових дерев на сталому наборі даних.

Ансамбль – це набір простих моделей, які разом дають деяку відповідь. Одна з переваг їх використання – декілька моделей, які намагаються спрогнозувати одну й ту ж змінну дадуть кращий результат, ніж одна модель. Техніки ансамблювання надалі розділяються на беггінг та бустинг.

Для беггінгу є сенс використовувати велику кількість дерев рішень із достатньою глибиною. В результаті класифікації фінальним результатом буде той клас, за який проголосувала більшість дерев ансамблю.

Random forest потребує багато ресурсів, але обмеження на глибину\кількість дерев значно знижує точність, тому важливим етапом підготовки моделі є вибір оптимальних значень її гіперпараметрів.

До основних параметрів моделі відносяться:

- кількість дерев;
- кількість ознак для відокремлення;
- мінімальна кількість об'єктів для яких виконується відокремлення;
- обмеження на кількість об'єктів в листах;
- максимальна глибина дерева;
- критерій відокремлення.

Що до кількості дерев, то чим їх більше, тим вище якість алгоритму на тренувальній підмножині, а на тестовій якість виходить на деяку асимптоту. Це дозволяє підібрати необхідну кількість дерев для кожної задачі окремо.

Кількість ознак для відокремлення – чим більша кількість цих ознак, тим більш однакові дерева отримуємо в результаті. За замовчанням він рівний  $\sqrt{n}$  в задачах класифікації та  $n/3$  в задачах регресії.

Мінімальна кількість об'єктів для яких виконується відокремлення – один з параметрів який можна залишати сталим – відповідає за рішення поділу отриманого листа.

Обмеження на кількість об'єктів в листах – схоже на попередній параметр, є обмеженням зверху на об'єм листа.

Максимальна глибина дерева – параметр що відповідає за якість моделі. Чим більша глибина, тим повільніше працює алгоритм. При використанні неглибоких дерев зміна параметрів об'єму листа не призводить до видимих змін. Через це рекомендується проводити тестування моделі на предмет виявлення найкращих параметрів.

Критерій відокремлення реалізований двома функціями MSE та MAE. Для класифікації реалізовані критерії gini та entropy.

Для підбору параметрів зазвичай використовується алгоритм перехресної перевірки по k блокам (k-fold Cross-Validation). Принцип його роботи полягає в тому що модель навчається на різних частинах тренувальної вибірки окремо і порівнюється результат.



Запускаючи кросс-валідацію ми отримуємо багато запусків моделі з різними параметрами, серед яких обирається та що показала найкращий результат.

В роботі використовувалась кросс-валідація для підбору найкращих параметрів, а потім виконувалось навчання і власне прогноз по кращій моделі.

### 1.7 Градієнтний бустинг

Коли ми намагаємося спрогнозувати цільову змінну за допомогою будь-якого алгоритма машинного навчання, головні причини відмінностей між реальними даними та прогнозом це шум, відхилення та зміщення. Ансамбль допомагає зменшити останні два фактори.

Ансамбль – це набір предикторів, які разом дають відповідь (наприклад, середнє по всіх). Причина чому так широко використовуються ансамблі – кілька предикторів, які намагаються отримати одну і ту ж змінну дадуть більш точний результат, ніж одиночний предиктор. Техніки ансамблювання згодом класифікуються в беггінг і бустинг.

Беггінг – проста техніка, в якій ми будуємо незалежні моделі і комбінуємо їх, використовуючи деяку модель усереднення (наприклад, зважене середнє, голосування більшості або нормальне середнє).

Зазвичай беруть випадкову вибірку даних для кожної моделі, так всі моделі трохи відрізняються одна від одної. Вибірка будується за моделлю вибору з поверненням. Через те що дана техніка використовує безліч некорелюючих моделей для побудови підсумкової моделі, це зменшує відхилення. Прикладом беггінга служить модель випадкового лісу (Random Forest, RF).

Бустинг – це техніка побудови ансамблів, в якій предиктори побудовані не незалежно, а послідовно.

Це техніка використовує ідею про те, що наступна модель буде вчитися на помилках попередньої. Вони мають нерівну ймовірність появи в наступних моделях, і частіше з'являться ті, що дають найбільшу помилку. Предиктори можуть бути обрані з широкого асортименту моделей, наприклад, дерева рішень, регресія, класифікатори тощо. Через те, що предиктори навчаються на помилках, скоєних попередніми, потрібно менше часу для того, щоб дістатися до правильної відповіді. Але тут також є свої нюанси – вибирати критерій зупинки треба з обережністю, інакше це може привести до перенавчання. В роботі було використано градієнтний бустинг.

Градієнтним бустингом називають техніку машинного навчання для задач класифікації та регресії, яка будує модель прогнозу в формі ансамблю слабких моделей, зазвичай дерев рішень.

Мета будь якого алгоритму навчання з учителем – визначити функцію витрат та мінімізувати її.

Якщо в якості функції втрат взяти середньоквадратичну помилку (MSE) (1.1):

$$L(y_i, y_i^p) = \text{MSE} = \sum (y_i - y_i^p)^2 \quad (1.1)$$

де  $y_i$  –реальне значення;

$y_i^p$  – прогнозоване;

$L(y_i, y_i^p)$  – Loss function.

Прогноз має будуватися таким чином щоб MSE була мінімальною. Використовуючи градієнтний спуск та оновлюючи прогнози, проводиться пошук значень на яких показник буде мінімальним (1.2, 1.3).

$$y_i^p = y_i^p + \alpha * \sigma \frac{\sum y_i - y_i^p}{\sigma * y_i^p}, \quad (1.2)$$

$$y_i^p = y_i^p - \alpha * 2 * \sum (y_i - y_i^p). \quad (1.3)$$

де  $\alpha$  – швидкість навчання.

В цілому, весь механізм градієнтного бустингу полягає в наступному:

- побудова простої моделі та аналіз помилок;
- відбір точок що не вписуються в просту модель;
- додавання моделей які обробляють точки з минулого пункту;
- збір усіх побудованих моделей та визначення ваги кожного предиктора.

## 1.8 Моделі глибокого машинного навчання

Сучасний термін «глибоке навчання» виходить за рамки нейро-біологічного погляду на моделі машинного навчання. В ньому закладено більш загальний принцип навчання декількох рівнів композиції що використовується у системах машинного навчання, не обов'язково влаштованих за принципом нейронів.

Основою сучасного машинного навчання були прості лінійні моделі на основі нейробіологічних аналогій. Вони приймали множину  $n$  вхідних значень  $x_1, \dots, x_n$  та асоціювали з ними вихід  $y$ . Під час тренування модель мала підібрати ваги  $w_1, \dots, w_n$  та обчислити вихід  $y$  вигляді (1.4).

$$f(x, w) = x_1 * w_1 + \dots + x_n * w_n \quad (1.4)$$

Ця перша хвиля нейронних мереж відома під назвою кібернетики.

Однак через великі обмеження на функціонал лінійних моделей (наприклад неможливість навчити модель функції XOR) багато хто заперечував можливість навчання на біологічних аналогіях.

Не зважаючи на це, нейробіологія дала надію на те що один алгоритм глибокого навчання зможе вирішити багато різних задач.

Інтерес до нейронних мереж був дуже високим в 1950–1960-х роках, коли були отримані перші приголомшливі результати. Після цього нейронні мережі трохи поступилися місцем іншим методам машинного навчання. Розробки нових систем продовжувались в інститутах в 1990-х роках, але особливого пожвавлення вони набули після 2010-х років.

### 1.9 Матеріали для навчання

Будь яке навчання мережі починається з даних. Під час шквального переходу на системи бухгалтерського обліку (славнозвісний 1С) багато представників бізнесу стикнулося з проблемами оцифровування даних. Всі бухгалтерські книги, відомості що до того осідали в товстих папках, почали плавно, часто за рахунок робочих годин працівників, перекочувувати в бази даних. Оскільки попит породжує пропозицію, почали з'являтися перші системи спрямовані на розпізнавання текстів, і відбулося збільшення об'ємів даних. Це спричинило ріст зацікавленості бізнесу в нових системах аналізу даних.

Згодом, коли у великих компаніях об'єм даних став хоч трохи задовольняти потребам нейромереж, почали розвиватися методи прогнозування. Спроби прогнозування продажів, робота з клієнтськими базами, класифікація та сегментація клієнтів, аналіз їх поведінки – алгоритми машинного навчання набували все більшої популярності та про них починали говорити.

На світових конференціях почали підніматися теми впливу алгоритмів на ведення бізнесу, приклади успішних кейсів IBM стали святим Граалем для переважної більшості представників малого та середнього бізнесу.

Другою проблемою стало те що можливості нових систем потребували збільшення потужностей підприємств, особливо з точки зору комп'ютерної периферії.

На цій хвилі хайпу почали з'являтися цілі підрозділи DataScience, які намагалися повторити відомі кейси на інших даних в рамках своїх компаній.

Все більш цікавим ставало використання нейромереж в аналізі продуктових чеків, закупок товарів, та прогнозуванні продажів.

Всі перераховані вище класи задач вирішувалися одним видом нейронних мереж – рекурентними нейронними мережами.

### 1.10 Рекурентні нейронні мережі

Рекурентні нейронні мережі це сімейство нейронних мереж для обробки послідовних даних, більшість з них також спроможні на обробку послідовностей змінної довжини [13].

В поставленій задачі вирішувалася проблема прогнозування послідовностей стабільної довжини, де кожен елемент послідовності – наповненість залу у кожен конкретний день минулого.

Розглянемо класичну форму динамічної системи (1.5):

$$s^{(t)} = f(s^{t-1}, \theta) \quad (1.5)$$

де  $s$  – стан системи.

Вираз є рекурентним, так як визначення  $s$  в момент часу  $t$  посиляється на те ж саме визначення в момент часу  $t - 1$ .

Для скінченного числа кроків  $t$  вираз можливо розвернути, використовуючи наведене вище твердження.

Наприклад (1.6):

$$s^{(3)} = f(s^{(2)}, \theta) = f(f(s^{(1)}, \theta), \theta) \quad (1.6)$$

Після такої розгортки отримаємо вираз що не містить рекурсії.

В якості іншого прикладу розглянемо динамічну систему що є керованою вхідним сигналом  $x(t)$  (1.7) :

$$s^{(t)} = f(s^{(s-1)}, x^{(t)}, \theta) \quad (1.7)$$

Тепер стан  $s$  містить в собі інформацію про всю минулу послідовність.

Коли рекурентну мережу навчають вирішувати задачу, в якій необхідно прогнозувати майбутнє по минулому, то мережа зазвичай навчається використовувати  $s(t)$  як набір послідовностей вхідних даних, що належать задачі. В загальному випадку в цьому наборі за необхідності втрачається частина інформації так як вона відображає послідовність змінної довжини ( $x(t)$ ,  $x(t-1)$ , ...  $x(1)$ ) на вектор фіксованої довжини  $s(t)$ . В залежності від критерія навчання, деякі аспекти минулої послідовності можуть мати перевагу над іншими [13].

Тепер лишилося розібратися з тим, як саме навчити нейромережу відображати вхідну послідовність на вихідну необов'язково такої ж довжини. Вхід такої РНМ часто називають «контекстом». Задача полягає в створенні представлення контексту  $C$ . Контекст  $C$  може бути вектором або послідовністю векторів що агрегує вхідну послідовність  $X = (x^{(1)}, x^{(2)}, \dots, x^{(n_x)})$  [13].

Наразі найбільш ефективними моделям послідовностей в практичних задачах є вентильні рекурентні нейронні системи. До них відноситься і довга коротокострокова пам'ять (англ. Long short-term memory – LSTM).

### 1.11 Моделі з довгою коротокостроковою пам'яттю

Модель LSTM є вентиляною рекурентною нейронною системою. Вентильні РНС засновані на ідеї прокладання шляхів крізь час, на яких похідні

не обнуляються і не зростають стрімко вверх [13]. В англomовній літературі ці ефекти часто називають *Vanishing and exploding gradients*.

У вентильних РНС ідею блоків пам'яті виконують ваги зв'язків, які можуть змінюватися на кожному кроці.

Блоки пам'яті дозволяють мережі накопичувати інформацію впродовж довгого періоду часу. Але інколи є сенс в тому щоб система поступово забувала старий стан при використанні недавньої інформації. Саме для програмування такого елемента «забування» використовуються вентильні РНС.

Принципiальна схема LSTM зображена на рисунку (рис. 1.2)

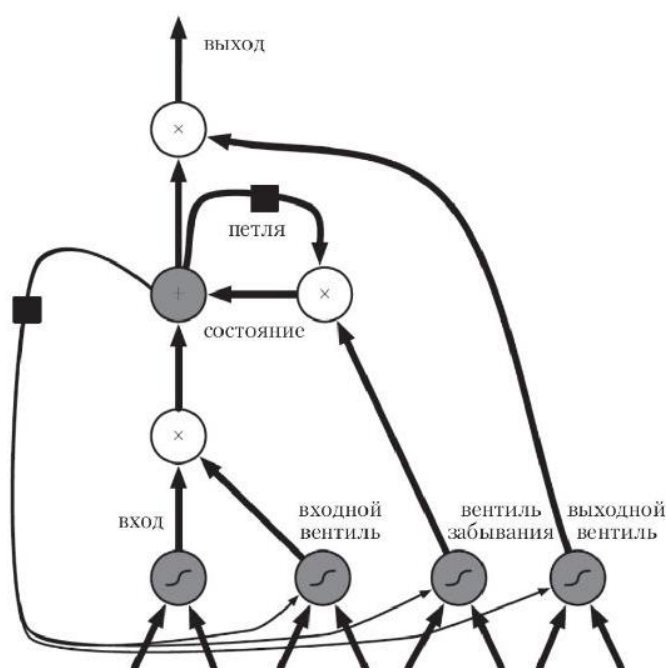


Рисунок 1.2 - Принципіальна схема LSTM

Замість блоку, який просто застосовує поелементну нелінійність до афінного перетворення входів та рекурентних блоків, в рекурентних LSTM мережах існують LSTM – комірки які містять у собі внутрішню рекурентність (петлю) в додаток до зовнішньої рекурентності РНС [13].

В кожній такої комірці є входи та виходи, такі ж як у звичайної мережі, але також там є додаткові параметри та система вентильних блоків що керує потоком

інформації. Найбільш важливим компонентом є блок стану  $s_i(t)$  з лінійною петлею. Вага цієї петлі керується за допомогою вентиляного блоку забування  $f_i(t)$  (для часового кроку  $t$  комірки  $i$ ) який присвоює цьому коефіцієнту ваги значення від 0 до 1 за допомогою сигмоїди (1.8):

$$f_i^{(t)} = \sigma(b_i^f + \sum_j U_{i,j}^f * x_j^{(t)} + \sum_j W_{i,j}^f s_j^{(t-1)}) \quad (1.8)$$

де  $x(t)$  – вхідний вектор на кроці  $t$ ,

$s(t)$  – вектор поточного прихованого шару, що містить виходи усіх LSTM комірок,

$b^f, U^f, W^f$  – зміщення, ваги входів та рекурентні ваги для вентилю забування.

Таким чином, внутрішній стан LSTM комірки оновлюється за наступною формулою, в якій присутня умовна вага петлі  $f$  (1.9).

$$h_i^{(t)} = f_i^{(t)} * h_i^{(t-1)} + g_i^{(t)} * \sigma(b_i + \sum_j U_{i,j} * x_j^{(t)} + \sum_j W_{i,j} s_j^{(t-1)}) \quad (1.9)$$

де  $b, U, W$  – зміщення, ваги входів та рекурентні ваги LSTM комірки.

Блок зовнішнього вхідного вентиля  $g_i^{(t)}$  розраховується аналогічно вентилю забування (з використанням сигмоїди для отримання значення від 0 до 1), але зі своїми параметрами (1.10).

$$g_i^{(t)} = \sigma(b_i^g + \sum_j U_{i,j}^g * x_j^{(t)} + \sum_j W_{i,j}^g s_j^{(t-1)}) \quad (1.10)$$

Історично доведено що LSTM мережі набагато легше навчаються довгостроковим залежностям, аніж прості рекурентні архітектури: спочатку проводилися тести на штучних наборах даних, спроектованих виключно для тестування здатності до навчання довгостроковим а згодом і на важких задачах обробки послідовностей, при вирішенні яких була досягнута якість що не була гіршою за найкращі результати .



## 1.12 Вимоги до програмного забезпечення

В роботі використано мову програмування Python, бібліотеки numpy, pandas, pymysql, json, requests, sklearn.

Програмний продукт розроблено в операційній системі Windows 10, але його використання не залежить від встановленої ОС.

Тестування проводилось в операційній системі Windows 10.

## Висновки до розділу 1

В даному розділі було розглянуто поняття математичної моделі, моделювання, наведено визначення понять беггінгу та бустингу моделей. Приведено класифікацію математичних моделей і методів, визначено до якого класу методів відноситься вибраний для роботи алгоритм.

Проаналізовано зовнішнє оточення проекту, виявлено відсутність аналогів продукту.

Розглянуто основні вимоги до програмної реалізації модульних систем.

Проаналізовано історію зростання попиту на нейронні мережі та основні перепони на шляху розробників.

Обрано інструмент для прогнозування серед великої кількості нейронних мереж, обґрунтовано його використання.

Приведено класифікацію математичних моделей і методів, визначено до якого класу методів відносяться вибрані для роботи алгоритми.

В результаті можна зробити висновок, що математичне моделювання це потужний інструмент для аналізу складних систем який можна використати для покращення програмного комплексу автоматизованого формування

розкладу сеансів для кінотеатрів на основі статистичних даних та даних схожих фільмів.

## Розділ 2 РОЗРОБКА НОВОГО МЕТОДУ

### 2.1 Первинний аналіз вхідних даних

Отримані дані можна класифікувати на групи за таким принципом:

- 1) безпосередні дані про фільм:
  - а) перелік топ акторів;
  - б) перелік режисерів;
  - в) показник популярності топу акторів;
  - г) показник популярності режисерів;
  - д) дані про сеанс фільму;
  - е) день прокату;
  - ж) час прокату;
  - з) кількість фільмів–прем'єр, що вийшли в той же день;
  - и) технологія прокату.
- 2) дані про схожі фільми:
  - а) середня наповненість залу;
  - б) середній показник популярності акторів;
  - в) середній показник популярності режисерів;
  - г) середній показник популярності фільмів.

В якості цільової змінної обрано показник наповненості фільму на сеансі.

Є доречним проговорити про кожну групу окремо.

### 2.2 Обґрунтування нового методу на формально-теоретичному рівні

Минулий підхід показав що модель, не знаючи інформації про інші фільми, заповнює зали найкращим фільмом. Були використані обмеження на кількість однакових сеансів в день, однак після тестування від цієї моделі вирішено було

відмовитись. Було недостатньо інформації про фільм, про паралельні фільми що виходять.

Крім того, було чітко видно що модель набагато краще прогнозувала перший тиждень, аніж наступні, тобто сенс днів прокату губився через велике різноманіття значень.

Враховуючи наведені вище фактори, було вирішено створити метод для автоматичного прогнозування розкладу сеансів для кінотеатрів, який включав би в себе певний комплекс моделей, та нівелював слабкі сторони попередньо створеної моделі.

Для того щоб покращити результат прогнозування, було вирішено взяти за цільову змінну не кількість людей на сеансі, а наповненість залу, так як нова величина є нормованою від 0 до 1, де 0 – пустий зал, а 1, відповідно, заповнений.

Крім того, враховуючи той факт що перша модель непогано прогнозувала перші дні прокату, вирішено було створити нову модель що б мала змогу бачити попередні продажі, це завдання було виконано за допомогою нейронних мереж з довгостроковою пам'яттю.

Під кінець, для коректної роботи був необхідний механізм для роботи з отриманими прогнозами, так як вручну розподіляти сеанси, навіть знаючи орієнтовну кількість відвідувачів – не легка справа. Для цього було вирішено задачу класифікації отриманих сеансів, де 1 – сеанс має бути включено до прокатного дня, а 0, відповідно – виключено.

З цих трьох частин було спроектовано метод, який дав змогу вирішити поставлену задачу автоматичного прогнозування розкладу сеансів для кінотеатрів (рис. 2.1).

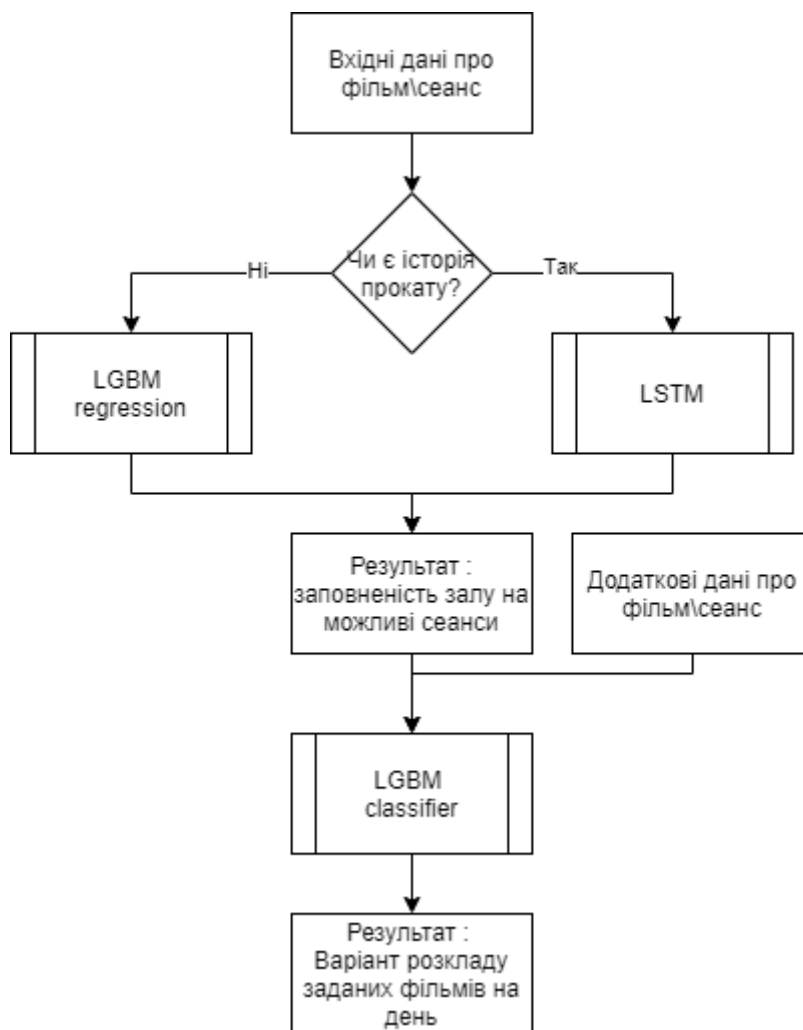


Рисунок 2.1 – Структура створеного методу для автоматичного прогнозування розкладу сеансів для кінотеатрів

Отже, новий метод (рис. 2.1) включає в себе наступні етапи:

1. Виконання процедури очистки та нормалізації вхідних даних.
2. Виконання аналізу історії прокату фільму (якщо є).
3. Прогнозування кількості людей на можливий сет сеансів.
4. Виконання аналізу додаткової інформації про сеанс.
5. Прогнозування можливого розкладу з можливого сету сеансів на 1 день.

### 2.3. Теоретичні оцінки параметрів розробленого методу

Безпосередні дані про фільм отримані за допомогою сторонніх сервісів (IMDB, TMDB, тощо) за допомогою написання скраперу даних. Скрапер аналізує контент сторінок по коду IMDB, який заповнюється оператором вручну при вводі фільма в базу кінотеатру. Зі сторінок ми отримуємо наступні дані:

- перелік топ акторів;
- перелік режисерів;
- показник популярності топу акторів;
- показник популярності режисерів.

Від жанрів та вікових обмежень вирішено було відмовитись через некоректний ввід цих показників в більшості фільмів. Під некоректним вводом особливо є сенс відміти що більшість фільмів 18+ позначено декількома мітками, такими як PG-13 та R, що аж ніяк не дає інформації про ту чи іншу вікову групу глядачів.

Перелік топ-акторів брався для отримання показнику популярності кожного актора на сайті. Наприклад такі актори як Джонні Депп(рис. 2.2) чи Кіану Рівз (рис. 2.3) мають значно вищі показники популярності аніж наприклад Джил Марія Джонс (рис. 2.4).

Body 17 Headers 0 Cookies

Pretty JSON Explorer Raw

```

1  {
2    "birthday": "1963-06-09",
3    "known_for_department": "Acting",
4    "deathday": null,
5    "id": 85,
6    "name": "Johnny Depp",
7    "also_known_as": [
8      "Джонни Депп",
9      "조니 데프",
10     "ジョニー・デップ",
11     "จอห์นนี่ เดปป์",
12     "強尼·戴普",
13     "Джонни Депп",
14     "約翰尼·德普",
15     "جوني ديب",
16     "Τζόνι Ντεπ",
17     "Τζον Κρίστοφερ Ντεπ II"
18   ],
19   "gender": 2,
20   "biography": "John Christopher \"Johnny\" Depp II (born June 9, 1963 in Kentucky) is an American actor",
21   "popularity": 30.947,
22   "place_of_birth": "Owensboro, Kentucky, USA ",
23   "profile_path": "/532tIC5wzmmR1j07oA83rTilvBx.jpg",
24   "adult": false,
25   "imdb_id": "nm0000136",
26   "homepage": null
27 }
28

```

Рисунок 2.2 – Профіль актора Джонні Депп

Response 200

Body 17 Headers 0 Cookies

Pretty JSON Explorer Raw

```

1  {
2    "birthday": "1964-09-02",
3    "known_for_department": "Acting",
4    "deathday": null,
5    "id": 6384,
6    "name": "Keanu Reeves",
7    "also_known_as": [
8      "Киану Ривз",
9      "키아누 리브스",
10     "キアヌ・リーブス",
11     "เคานู รีฟส์",
12     "基努·里维斯",
13     "קיאנו ריבס",
14     "Keanu Charles Reeves",
15     "Κιάνου Ριβς",
16     "Κιάνου Τσαρλς Ριβς",
17     "Киану Ривз",
18     "کیانو ریوز"
19   ],
20   "gender": 2,
21   "biography": "Keanu Charles Reeves is a Canadian actor. Reeves is known for his roles in Bill & Ted's",
22   "popularity": 25.171,
23   "place_of_birth": "Beirut, Lebanon",
24   "profile_path": "/rRdru6REr9i3WtHv2mntpcgxnoY.jpg",
25   "adult": false,
26   "imdb_id": "nm0000206",
27   "homepage": null
28 }

```

Рисунок 2.3 – Профіль актора Кіану Рівз

## Response 200



Рисунок 2.4 – Профіль актора Джилл Марі Джонс

Показник популярності акторів є важливим в аналізі фільму. Чим більше відомих людей там знімається, тим з більшою ймовірністю обізнана людина піде на цей фільм.

Також, оскільки цей показник не особливо змінюється в часі та не напряму залежить від новизни фільму, в якому знімався актор, то ці дані можна брати в розрахунки для моделі за останні декілька років.

З режисерами схожа ситуація, зазвичай фільми приквели та сиквели знімаються одним і тим же режисером, отже люди що були на першому фільмі і він їм сподобався, підуть і на другий. Перша думка була взяти режисера як фічу, однак через велику розрідженість матриці фільм-режисер довелося взяти лише популярність режисера. Чим більше у нього успішних фільмів – тим вище рейтинг.

Через це було узгоджено використання у роботі двох важливих показників – середній рейтинг акторів фільму та середній рейтинг режисерів фільму. Ці дані описують якість фільму нормованими показниками.



Тепер переходимо до даних про сеанс. Це також важлива частина даних, але вона вже відноситься не до фільму безпосередньо, а саме до об'єкту що містить у собі фільм. Сеанс оперує такими показниками як час, день прокату фільму, та технологія в якій катається фільм.

Від часу залежить кількість людей що зможе прийти, дитячі фільми ставлять переважно на ранковий час, бестселери – на прайм–тайм, який кожен кінотеатр виділяє по своєму в залежності від розташування комплексу або ТРЦ в якому знаходиться кінотеатр.

Від дня прокату залежить ціна на квитки, тип залів в яких буде кататися фільм, кількість сеансів фільму, тобто те з чого вибирає глядач, ідучи на фільм.

Технологія прокату також грає важливу роль в оцінці кількості людей, так як зали з різними технологіями мають різну кількість місць, в 4Dх наприклад їх менше ніж в інших технологіях. Ці показники залежать від кожного кінотеатру.

На відміну від попередньої моделі, було вирішено додати інформацію про те яка кількість прем'єр виходить на тижні прокату, з чого обирає глядач, з великої кількості новинок, чи, як зараз, в часи пандемії, з декількох фільмів.

Отже до групи даних про сеанс фільму ми віднесли наступні показники:

- день прокату;
- час прокату;
- кількість фільмів–прем'єр, що вийшли в той же день;
- технологія прокату.

Під параметром день прокату мається на увазі кількість днів від початку прокату.

Третьою групою показників ми обрали дані про схожі фільми. Для того щоб оцінити якість фільму, оператор зазвичай сам аналізував які були схожі фільми в прокаті кожного конкретного кінотеатру\мережі кінотеатрів. Після цього, він дивився на заповненість залів, аналізував якість схожих фільмів, і вручну ставив необхідні сеанси.

Всі ці дії ми автоматизували за допомогою алгоритму пошуку схожих фільмів, і використовуючи працюючий додаток, змогли отримати дані

заповненості.

Крім того, за допомогою описаного вище скрапера, ми отримали змогу дивитись рейтинги акторів та режисерів схожих фільмів, отже, порівнювати та аналізувати знайдені фільми.

Дані про схожі фільми були введені як суттєве покращення минулої моделі, і до них належать наступні показники:

- середня наповненість залу;
- середній показник популярності акторів;
- середній показник популярності режисерів;
- середній показник популярності фільму.

Дані що до часу прокату після детального аналізу було вирішено поділити на чотирьох-годинні проміжки (рис. 2.5)

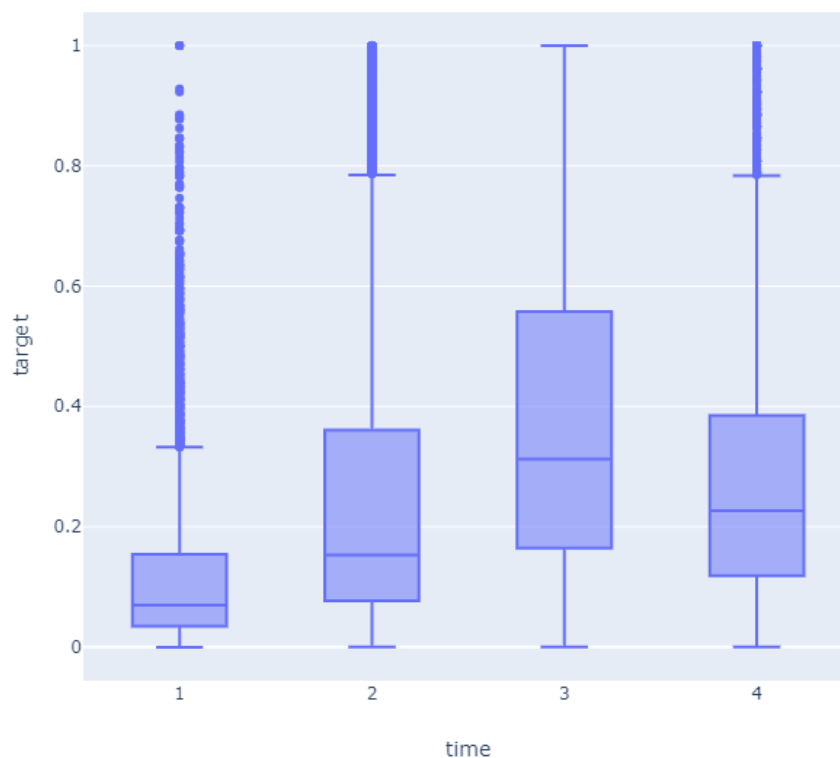


Рисунок 2.5 – залежність відвідувань від часу сеансу

На діаграмі (рис. 2.5) взято чотирьох–годинні проміжки для кожного сеансу наступним чином (2.1):

$$\begin{cases} 1, 8:00 - 12:00 \\ 2, 12:00 - 16:00 \\ 3, 16:00 - 20:00 \\ 4, 20:00 - 24:00 \\ 5, \text{else} \end{cases} \quad (2.1)$$

З приводу використання наповненості залів по схожим фільмам, був проведений аналіз на деякому відсотку фільмів, і легко побачити що цей показник непогано описує можливу кількість глядачів в кожному часовому проміжку (рис. 2.6).

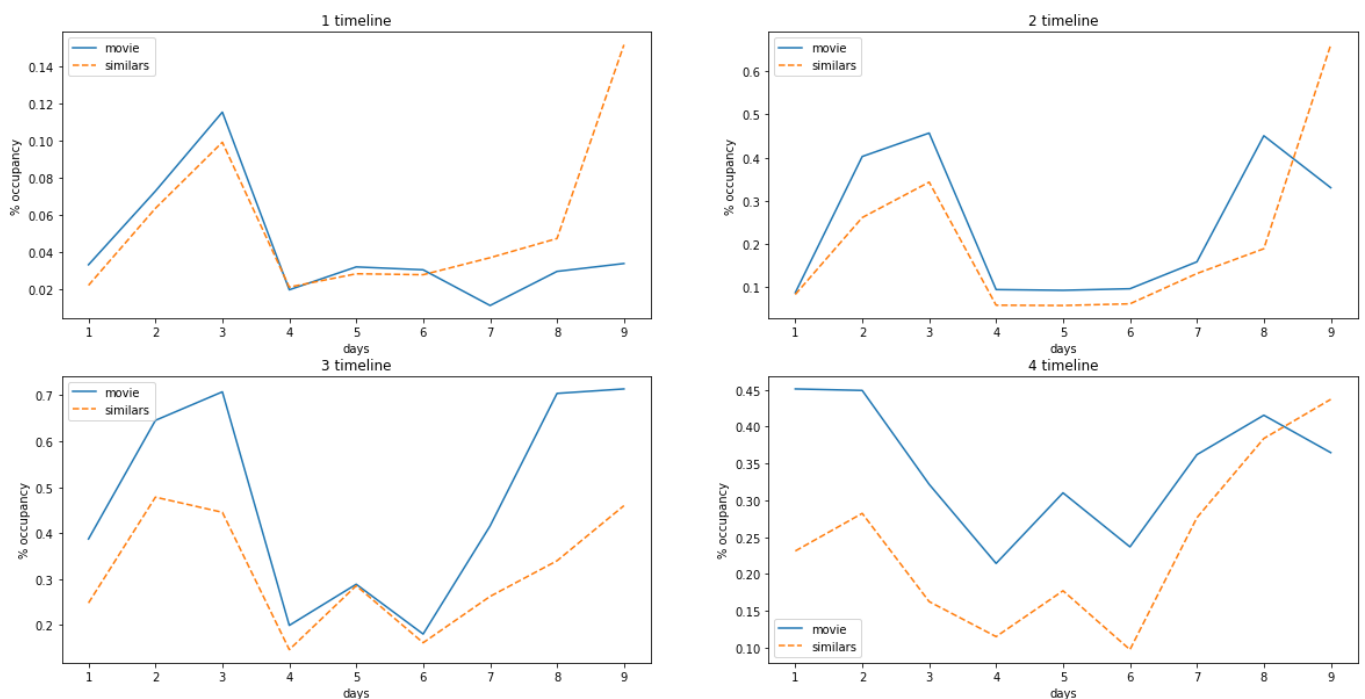


Рисунок 2.6 – Приклад опису можливої кількості глядачів за допомогою наповненості залів по схожим фільмам

Також, аналізуючи вплив кожного показника на цільову змінну target, бачимо що найбільш незначний вплив має саме показник кількості глядацьких

голосів за схожі фільми, а найбільш значний – як раз наповненість залів по схожим фільмам. Сюди не включені показники день, технологія та часова рамка, так як ці змінні є категоріальними.

Для більшого розуміння зв'язку параметрів між собою у роботі наведено кореляційну матрицю (рис. 2.7)

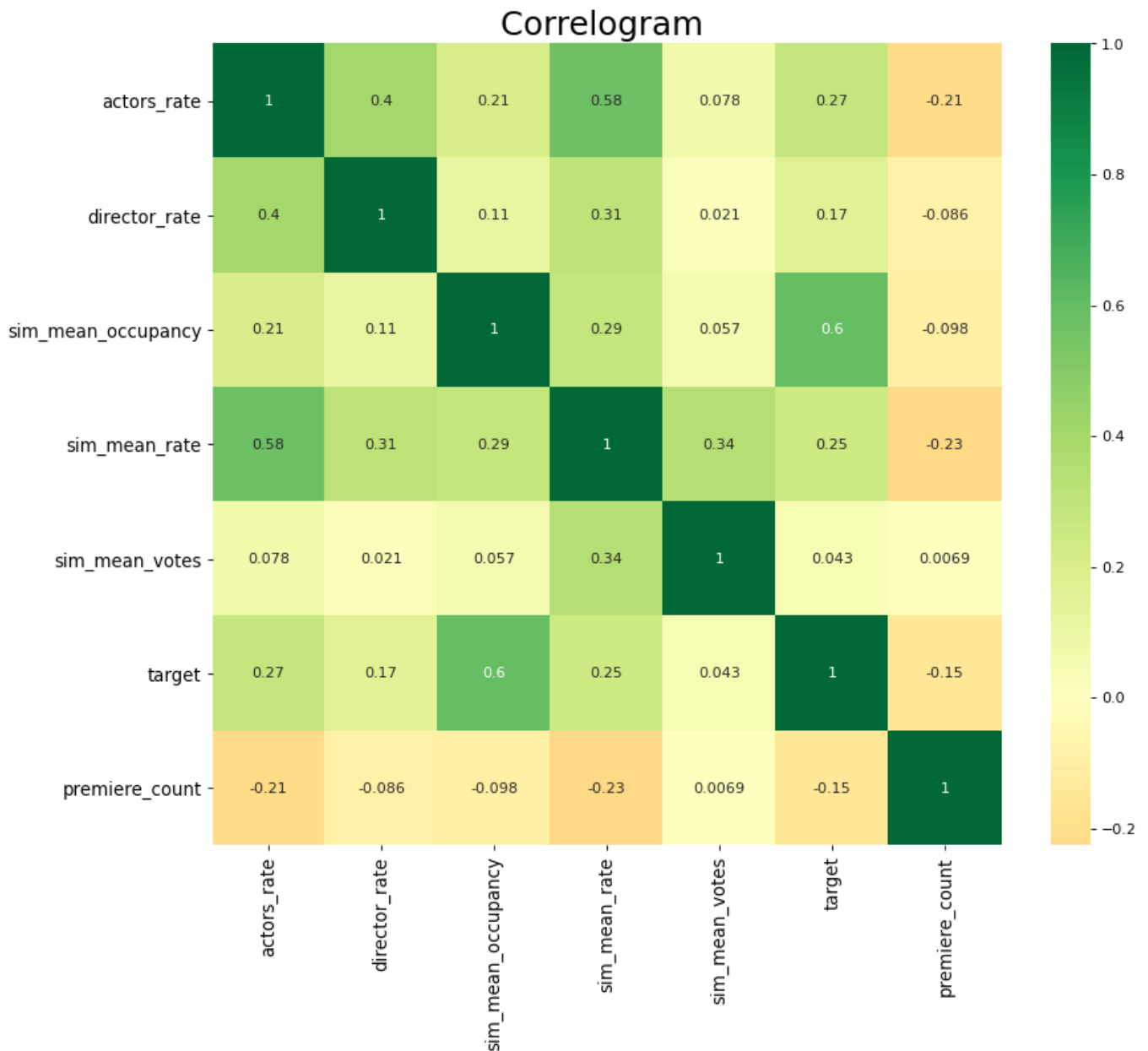


Рисунок 2.7 – Кореляційна матриця показників що будуть включатися до майбутньої моделі

Що до другої моделі, то її архітектура є доволі простою (рис. 2.8).

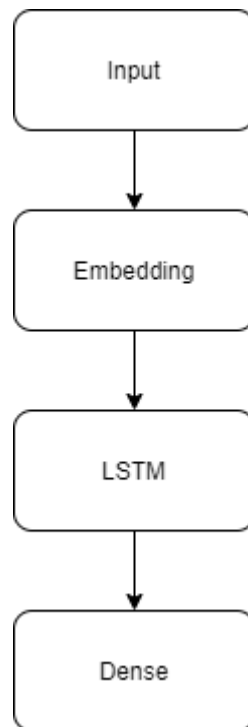


Рисунок 2.8 – Архітектура моделі для фільмів з історією

Ідея такої архітектури полягає в тому що спершу за допомогою шару Embedding система отримує інформацію про попередню заповненість сеансів (по суті, це інформація про тренд). Шар LSTM прораховує можливий варіант послідовності, і нарешті Dense який повертає саме наступні числа – заповненість залів на схожі сеанси.

Дані для входу в модель ми готуємо за допомогою вбудованих функцій мови python.

На вхід третьої моделі (знов LightGBM) подаються прогнози перших двох моделей, які в таблиці названо predict (рис. 2.9).

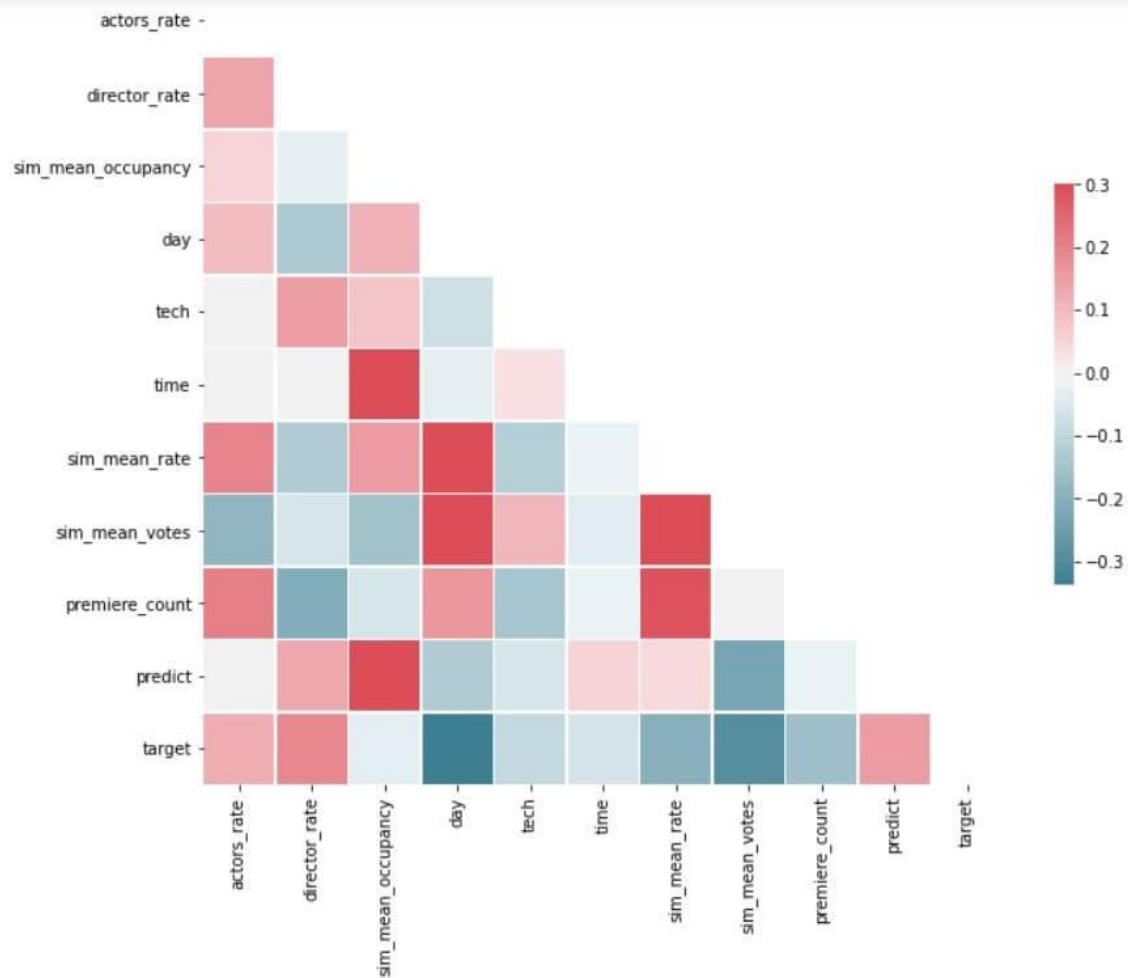


Рисунок 2.9 – кореляційна матриця для фіч третьої моделі

Новизною методу є те, що ми намагаємося навчити систему саму обирати кращі сеанси, базуючись на знаннях людини. На вхід ми подаємо всі можливі комбінації сеансів, а на виході для навчальної вибірки помічаємо 1 як сеанси що відбулися насправді, а 0 – як ті що не відбулися. Розклади справжніх сеансів бралися з бази даних кінотеатрів.

Звісно що така модель все одно буде потребувати ручних коригувань, як мінімум тому що є поняття оренди залів, в яких певний час не проводяться фільми, але в цілому вона має значно спростити формування сеансів на кожен день.

## Висновки до розділу 2

В другому розділі було виконано детальний огляд вхідних даних, попередній аналіз можливих ознак для побудови комплексу моделей для прогнозу очікуваної наповненості залу на сеанс та розподілення сеансів на день, а також спроектовано метод, який дав змогу вирішити поставлену задачу автоматичного прогнозування розкладу сеансів для кінотеатрів.

## Розділ 3 ОПИС РОЗРОБЛЕНОГО ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ

### 3.1 Структура програмного комплексу

Для побудови математичної моделі необхідно було проаналізувати недоліки існуючих систем та включити в модель варіанти їх вирішення.

Основними недоліками існуючих систем є ручне формування розкладу та відсутність бази фільмів з обновлюваною інформацією.

Щоб вирішити проблему нестачі інформації по прокату фільму для створення розкладу, було вирішено розробити програму яка б шукала схожі фільми по наданій інформації з IMDB.

Для реалізації параметра вибору сеансів було вирішено використати модель LightGBM Regressor для прогнозу наповненості залу для сеансів фільмів які каталися менше 14 днів, та нейронну мережу на основі LSTM для прогнозу наповненості залу сеансів фільмів які каталися більше 14 днів.

В свою чергу, для створення автоматичної системи вибору найкращих сеансів було вирішено використати модель LightGBM Classifier. Модель системи було поділено на три частини (рис. 3.1).

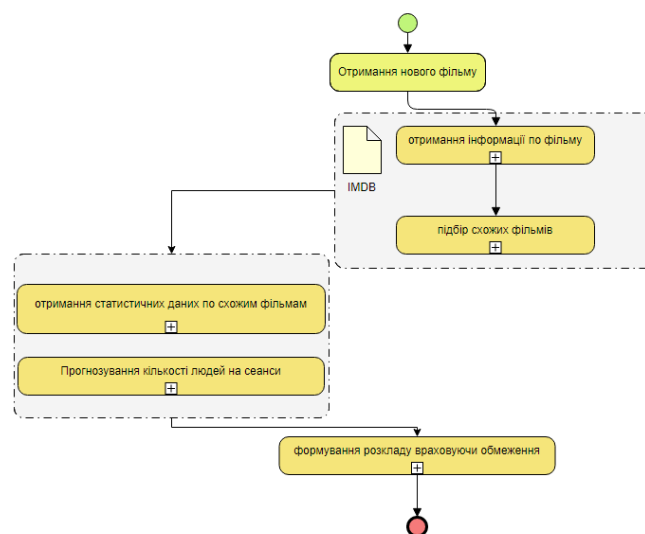


Рисунок 3.1 – Діаграма створеної математичної моделі



Перша частина відповідає за збір даних, їх обробку, та підбір схожих фільмів.

Друга частина працює напряду з даними користувача кожного конкретного кінотеатру, завантажує та обробляє їх, та на їх основі робить передбачення для можливої кількості людей на сеанс. Цим вирішується проблема недостатньої кількості інформації про фільм що виходить – підбираються схожі фільми про які багато історії, та на їх основі навчається модель.

Третя частина розроблена для того щоб врахувати обмеження на кількість сеансів в кожному залі, враховуючи години роботи кожного кінотеатру.

В цілому ці три частини склали програмний комплекс, який було вирішено реалізувати мовою програмування Python, та випробувати на реальних даних одної з мереж кінотеатрів України.

### 3.2 Опис розробки програмного комплексу

На першому етапі було виконано збір даних. За основу взяли базу IMDB, та його дзеркало – арі TMDb (табл. 3.1).

Таблиця 3.1 – Приклад отриманих даних

Назва	Приклад значення	Формат
MovieName	Alita: Battle Angel	string
MovieRating	7.6	float
Genres	Action, Adventure, Sci-Fi	Array of string
Rate	PG-13	string
Tags	robot, future, anime	array of string
Country	USA, Japan	array of string
Actors	Rosa Salazar, Christoph Waltz, Jennifer Connelly	array of string

Надалі цей механізм отримання даних було автоматизовано до стану онлайн отримання даних через мікросервіс.

Механізм його роботи описується наступним алгоритмом (рис. 3.2) :

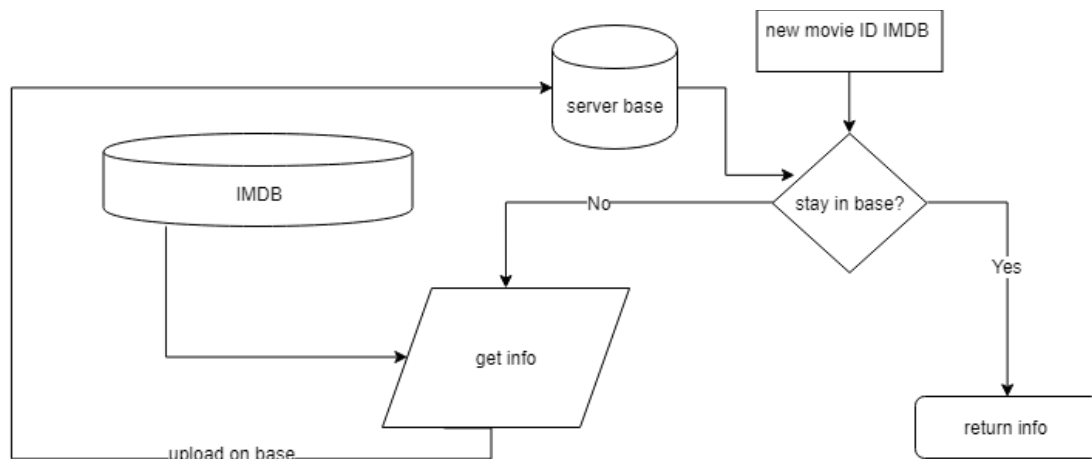


Рисунок 3.2 – Механізм роботи мікросервісу для отримання даних

Користувач подає запит у вигляді POST запиту наступного формату (3.1, 3.2):

[http://\[host\]:\[port\]/statistics/current](http://[host]:[port]/statistics/current) (3.1)

body: {"imdbId": "tt0437086"} (3.2)

Запит надходить до сервера, зчитується інформація з поля `imdbId`. Виконується перевірка, чи є фільм в локальній базі сервера, якщо так то інформація по фільму повертається з локальної бази без підключення до бази IMDB.

Якщо ж фільм відсутній у базі, то запускається виконання скрапінгу по сторінці на сайті IMDB, інформація надходить у локальну базу, зберігається там і віддає отриману інформацію користувачу.

Також там є окремі команди що відповідають за оновлення бази, отримання усіх фільмів, тощо. Автоматично оновлення завантажуються в момент найнижчої активності системи – вночі.

Надалі інформація по новому фільму завантажувється в другу частину алгоритму, яка здійснює підбір схожих фільмів.

Принцип її роботи наступний (рис. 3.3):

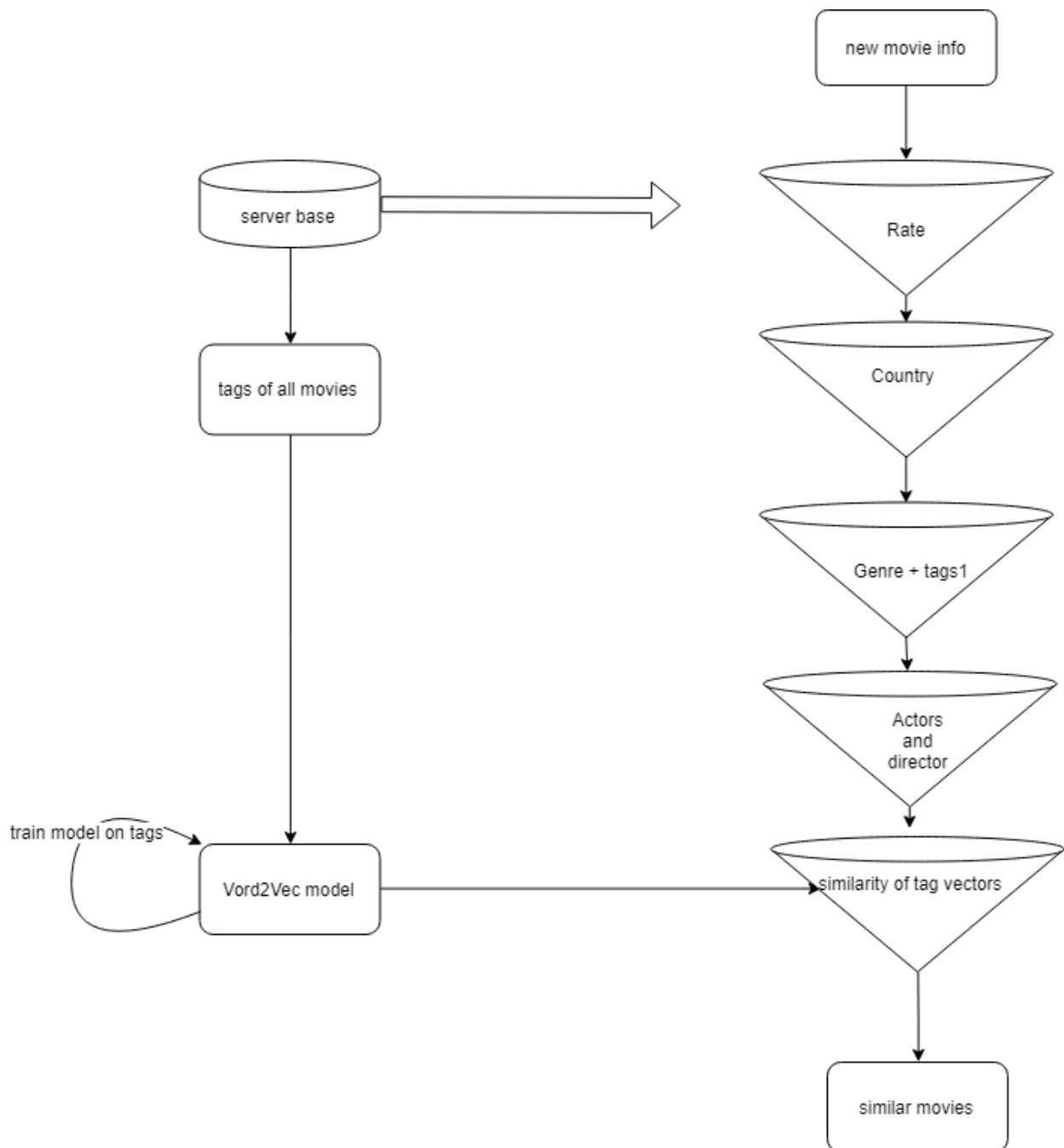


Рисунок 3.3 – Механізм підбору схожих фільмів

Сама модель підбору була перевірена на багатьох фільмах та експертна оцінка що до результатів її роботи є хорошою.

Спершу ми відбираємо фільми по віковим обмеженням на 2 категорії: [G, PG] та [PG-13, R], відділяючи таким чином дитячі мультики від фільмів.

Далі сортуємо за країною–виробником, відділяючи в окрему категорію українські та російські фільми (рис. 3.4). Крім того, в деяких країнах є своє, авторське кіно, цей підбір вирішує проблему таких рекомендацій.

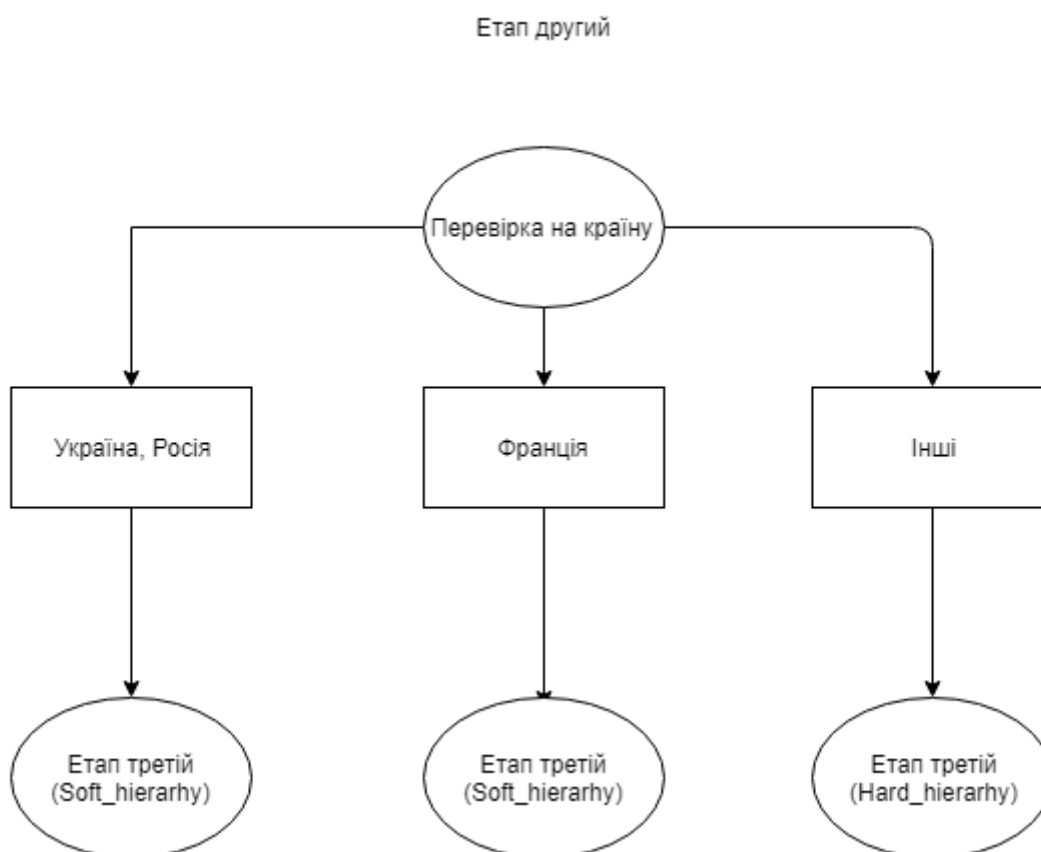
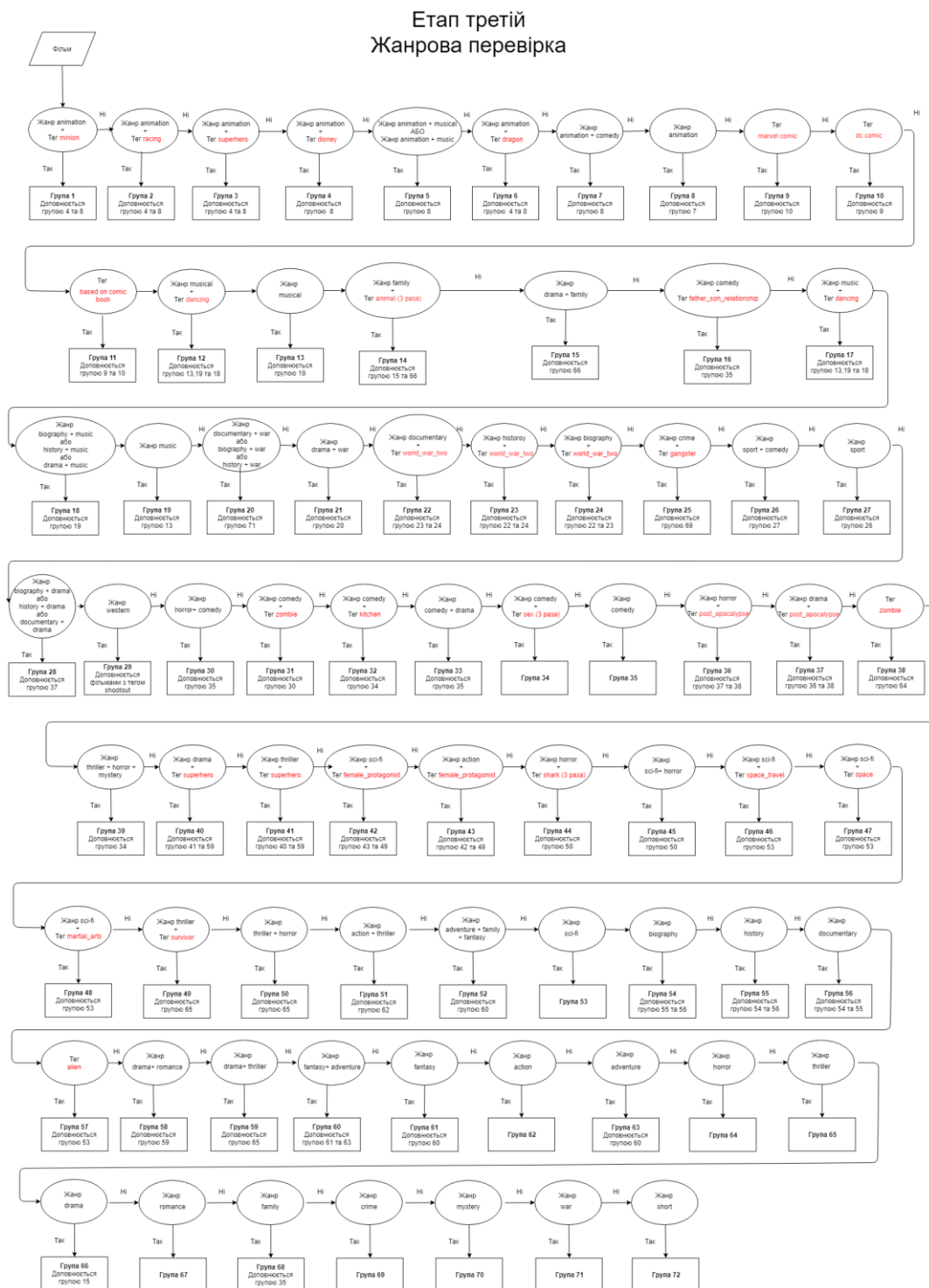


Рисунок 3.4 – сортування за країною–виробником

Третій пункт – сортування у зв’язці «Жанр + Тег». Теги, які вирізняють цілі розділи фільмів, такі як Space, Postapokalypsis, Alien, Timeloop, тощо. В об’єднанні з тегами вони дають хороший результат (рис. 3.5, 3.6).



Потім в кожній з груп неймережа виділяє топ 10 найбільш релевантних фільмів аналізуючи теги. Якщо в групі не вистачає фільмів, доповнити групу можна фільмами з іншої будь-якої групи

Рисунок 3.5 – сортування у зв'язці жанр + тег

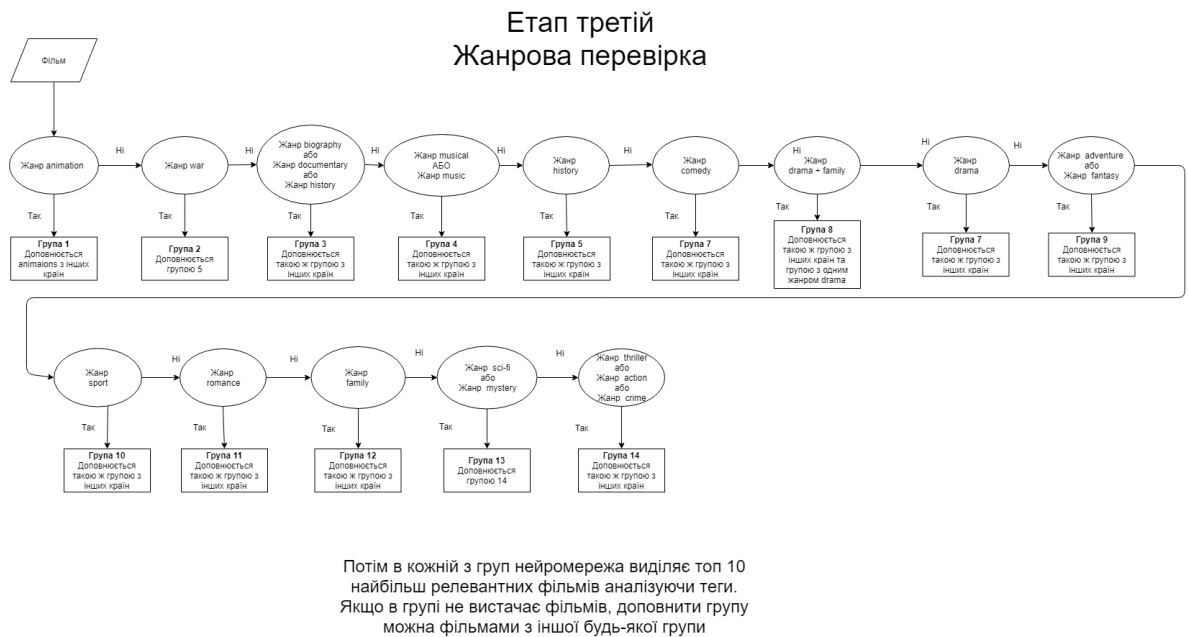


Рисунок 3.6 – жанрова перевірка

Четвертим пунктом є градація у отриманих списках фільмів. Також за допомогою цієї частини знаходимо сиквели та приквели до фільмів. Зазвичай в них однакові режисери та здебільшого співпадає акторський склад. Ця градація присвоює оцінки (в залежності від кількості збігів), та список сортується вже за цією оцінкою.

Останнім етапом є використання моделі Word2Vec, що являє собою інструмент для аналізу семантики природніх мов. Його технологія базується на векторному представленні слів. Модель вчиться в режимі CBOW, намагаючись передбачити слово, виходячи з його контексту. На виході отримуємо координатні представлення векторів–слів, для яких обчислюємо семантичну відстань між словами.

За допомогою цього показника відбувається сортування всередині списку отриманих фільмів – щоб виловити найбільш схожі за тегамі фільми в обраній категорії.

Надалі з кожної відсортованої категорії обирається топ–10 найкращих фільмів. Ця частина програми також виконана як окремий мікросервіс (Додаток А рис. 1, рис. 2).

Наступна частина програми – програмний комплекс для формування розкладу (рис. 3.7).

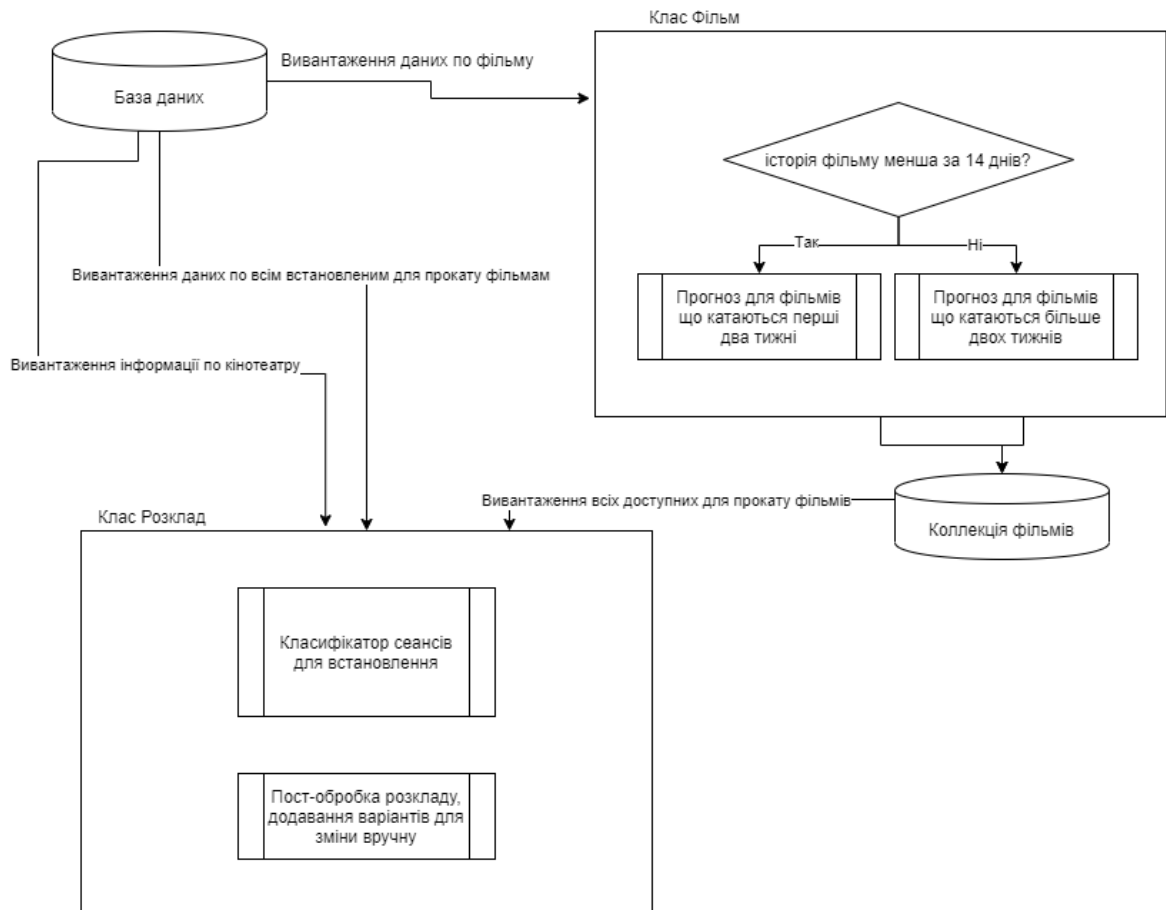


Рисунок 3.7 – Структура програмного комплексу для формування розкладу

Весь програмний комплекс доречно розбити на дві великі частини, окремо розділивши аналіз даних та прогнози та окремо - їх обробку [14], [15], [16].

Перша частина реалізована за допомогою класу Фільм (рис. 3.8)

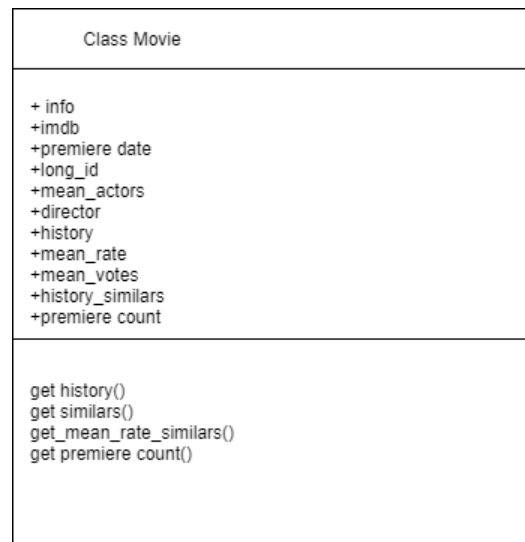


Рисунок 3.8 – Структура класу Фільм

Об'єкт класу Фільм є агрегатором даних для кожного фільму в базі, аналогічно з логікою, представленою в роботі для пацієнтів [14], [15], [16]. В ньому реалізовано функції отримання історії прокату з бази даних, вивантаження схожих фільмів з програмного комплексу для схожих фільмів, що був реалізований в минулій роботі, аналіз рейтингів схожих фільмів, і також підрахунок кількості фільмів що виходять в той же день.

Для зручності користування, було створено клас Сеанс, в якому реалізовано саме роботу з створеними моделями прогнозу наповненості залу (рис. 3.9)

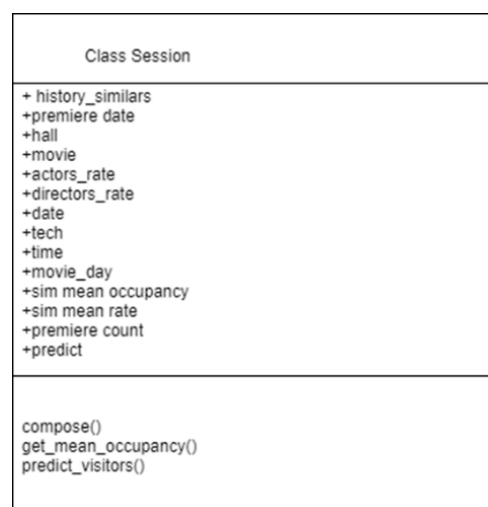


Рисунок 3.9 – Структура класу Сеанс



В об'єкті класу сеанс робота ведеться безпосередньо з моделлю за допомогою функції `predict_visitors()`. В залежності від кількості записів в історії фільму, використовуються модель для прогнозу перших двох тижнів, або модель для прогнозу тижнів, починаючи з третього.

Для роботи з сеансами, на які була зпрогнозована наповненість залу, створено клас Розклад (рис. 3.10).

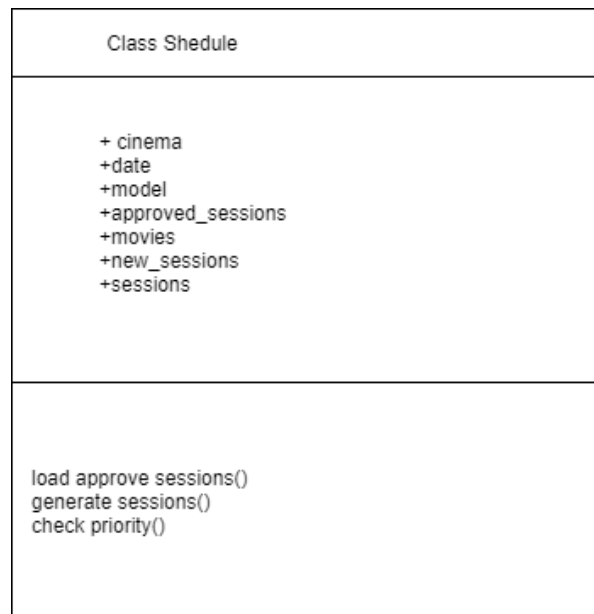


Рисунок 3.10– Структура класу Розклад

В об'єкті класу Розклад реалізовано функції завантаження сеансів, що були встановлені раніше.

Функція `generate sessions` відповідає за завантаження та генерацію усіх можливих сеансів, які ще не встановлено.

Функція `check priority` виконує пост-обробку отриманих сеансів, групує їх за залами та годинами, таким чином формуючи додаткові сеанси, які можна за бажанням замінювати вручну в створеному раніше інтерфейсі.

Спершу дані прокату фільмів формуються користувачем в візуальному інтерфейсі.

Після введення коду IMDb та натискання кнопки «Заповнити» формується запит до мікросервісу отримання даних, та інші поля заповнюються автоматично (Додаток А рис. 3).

У випадку якщо щось введено некоректно або не відповідає прокатній картці фільму, користувач може внести зміни та зберегти інформацію в програмі.

Далі з прокатної картки переноситься інформація що до обмежень прокату – в яких технологіях, в який період часу та скільки мінімально разів має кататися фільм щоденно (рис А.4).

Надалі користувач вибирає необхідні фільми, кінотеатр, та кількість днів для формування розкладу. Ці дані записуються в файл json, і подаються на вхід програми.

Далі з бази даних користувача завантажується існуюча інформація про фільм. Це має бути дані про прем'єру фільму та ідентифікатор даного фільму. Словник значень з ключами Прем'єра та фільм подається на вхід програмного комплексу.

По кожному фільму зі списку завантажуюмо статистичні дані з Базы продажів. Інформація що отримується, виглядає наступним чином (табл. 3.2):

Таблиця 3.2 – Інформація з «Базы продажів»

Змінна	Значення
CountID	К-ть куплених білетів
Movie_code	Внутрішній код фільму
Data	Дата сеансу
Time	Час сеансу
TehCode	Технологія
Cinema	Кінотеатр
Hall	Зал
Seats	К-ть місць
sessionID	Номер сеансу

Після цього проводиться обробка даних. Час сеансу розбивається відповідно (3.3):

$$\begin{cases} 1, 8:00 \leq t \leq 12:00 \\ 2, 12:00 \leq t \leq 16:00 \\ 3, 16:00 \leq t \leq 20:00 \\ 4, 20:00 \leq t \leq 24:00 \\ 5, \text{else} \end{cases} \quad (3.3)$$

Кожна цифра – чотирьох-годинний проміжок, куди можна поставити фільм. Таких проміжків 5, в подальшому проміжки від 1 до 2 – це ранок, а від 3 до 5 – вечір. Це необхідно для розбиття часового інтервалу трансляції на денні та вечірні сеанси.

Далі програмно створюються нові змінні:

- кількість днів з релізу фільму;
- час сеансу;
- середній показник заповненості схожих фільмів на такий же день прокату;
- середнє рейтингу акторів;
- середнє рейтингу режисерів;
- середнє рейтингу схожих фільмів;
- технологія прокату;
- кількість прем'єр що виходять одночасно;
- історія прокату фільму (якщо є).

В залежності від наявності історії прокату фільму, використовується модель LightGBM Regressor або нейронна мережа на основі LSTM.

Запуск цього механізму відбувається всередині класу Розклад, куди надходить інформація про фільми та обмеження. Після створення всіх доступних сеансів та прогнозу наповненості залу на них за допомогою попередніх кроків, дані завантажуються в модель для вибору найкращих сеансів.

Для навчання цієї моделі використовувалися готові розклади обраного кінотеатру, в яких одиницями були помічені усі встановлені сеанси, а нулями – всі генеровані.

Нижче представлений формат даних на вхід останньої моделі (табл. 3.3).

Таблиця 3.3 – Формат даних на вхід моделі вибору сеансів

Параметр	Пояснення
actors_rate	Середнє рейтингу акторів
director_rate	Середнє рейтингу режисерів
sim_mean_occupancy	Середнє наповненості схожих фільмів
day	День прокату
tech	Технологія
time	Час прокату [1–5]
sim_mean_rate	Середнє рейтингу схожих фільмів
sim_mean_votes	Середнє глядацьких вподобань схожих фільмів
premiere_count	Кількість прем'єр що виходять одночасно
predict_occupancy	Зпрогнозоване попередньою моделлю значення наповненості залу
target	0 або 1 – в залежності від того чи є сеанс в запланованих.

### Висновки до розділу 3

Було створено програмний продукт що дав змогу отримати необхідну інформацію по фільму, підібрати схожі фільми з минулих прокатів, по ним навчити модель прогнозувати наповненість залів в залежності від фільму та часу, а також сформувані базову версію розкладу сеансів у обраному кінотеатрі на тиждень.

Також в програмному продукті передбачена можливість вносити ручні правки фахівцями області, так як в деяких ситуаціях необхідне втручання людини для коригування даних про фільми.

## Розділ 4 ОПИС ВИПРОБУВАНЬ ПРОДУКТУ

### 4.1 Сервіс для підбору схожих фільмів

Тестування сервісу для підбору схожих фільмів проводилось за рахунок кінцевих користувачів. Продукт розміщено на сервері замовника (рис. 4.1, 4.2).

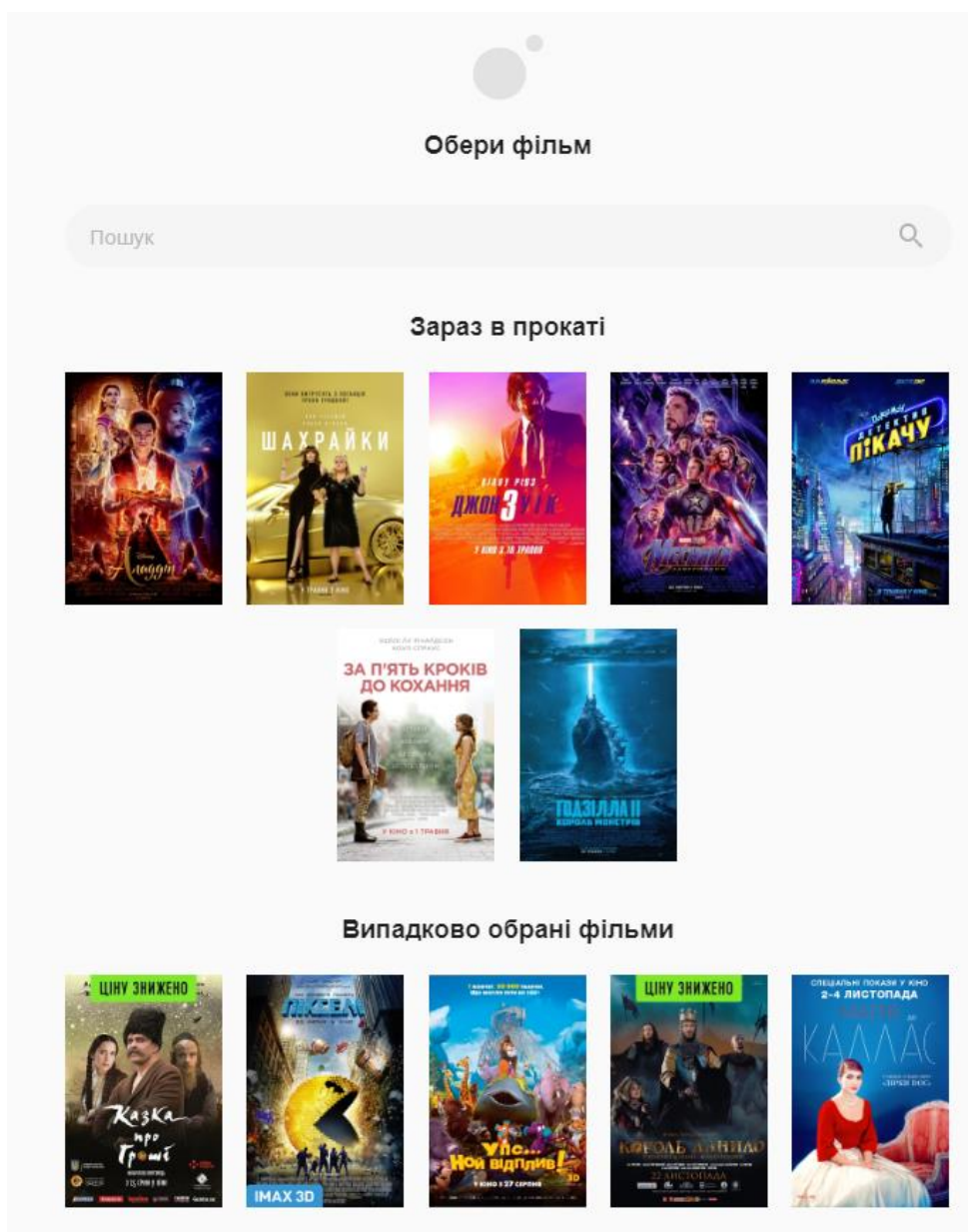


Рисунок 4.1 – Вхідна сторінка сервісу для підбору схожих фільмів

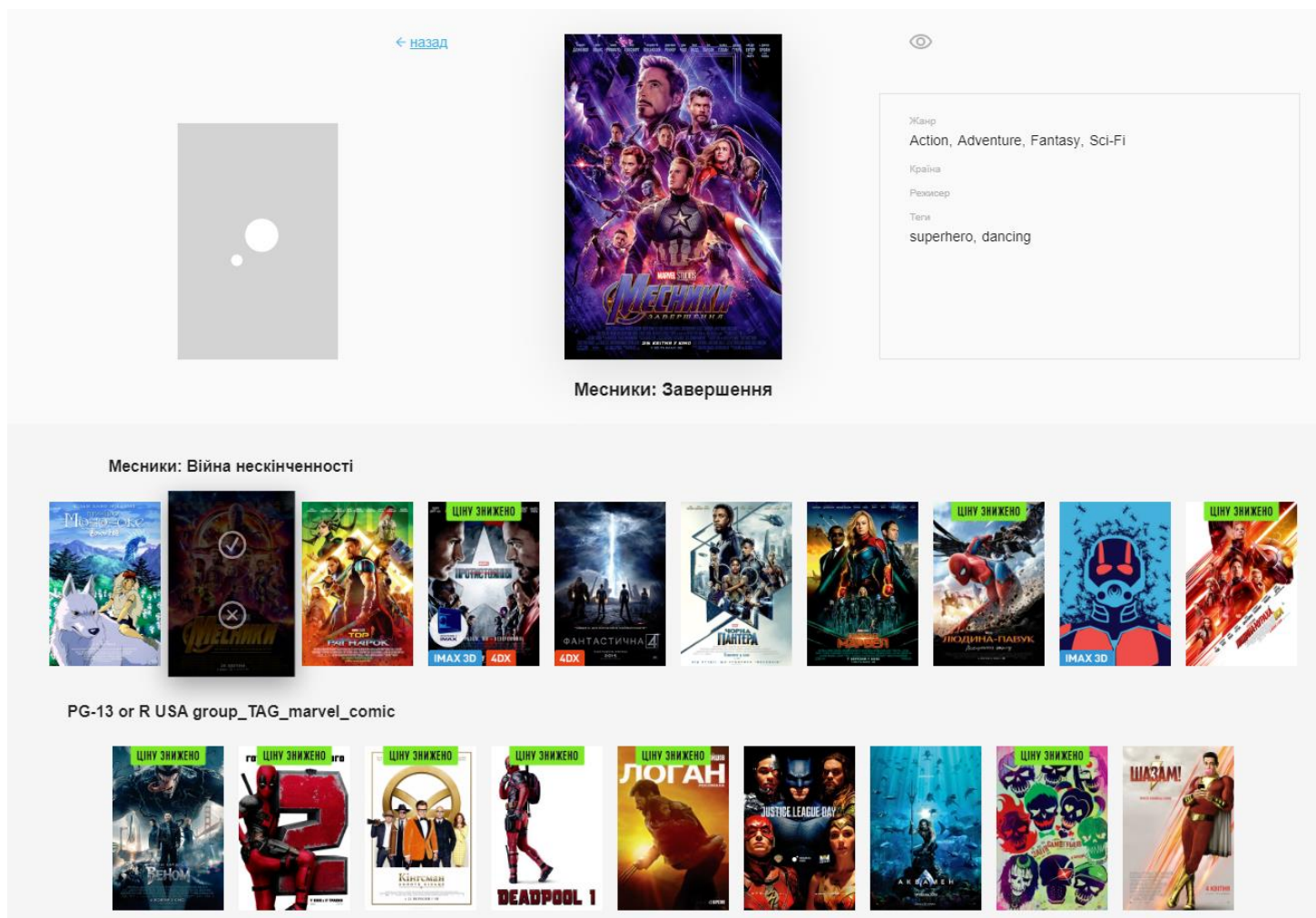


Рисунок 4.2 – вікно підбору схожих фільмів

В сервісі є можливість редагувати теги, жанри, продивляться коротку інформацію по фільму, також користувач може додавати схожі, на його думку, фільми. Всі зміни зберігаються в базі, звідки відбувається завантаження інформації на тестовий та справжній сайти.

Крім того при наведенні курсору на стручку схожих фільмів, знизу з'являється пояснення, за якими критеріями була збіжність.

## 4.2 Програмний комплекс для формування розкладу

Тестування створеного програмного продукту проводилось в три етапи:

- перевірка точності роботи алгоритму підбору схожих фільмів;
- коректність роботи моделі що прогнозує заповненість залу;
- коректність роботи моделі що обирає сеанси для розкладу.

Першим етапом була перевірка точності роботи алгоритму підбору схожих фільмів. Тестування виконувалось експертами зі сторони кінцевого користувача.

Результат цієї перевірки підтвердив достатню коректність результатів підбору.

Алгоритм користування цим модулем доволі простий:

- перейти на вкладку з вікном вибору фільмів для аналізу (Додаток А рис. 1);
- натиснути на фільм з зони «зараз в прокаті», або з «випадково обрані фільми», або зі списку пошуку.

Після цього внизу буде топ-10 фільмів, схожих на обраний фільм.

Для того щоб подивитися який відсоток їх схожості, потрібно натиснути на іконку ока в верхньому правому куті сторінки.

Ще реалізована можливість подивитися за якими параметрами вони схожі – для цього потрібно навести курсор на схожий фільм.

Надалі потрібно було перевірити коректність роботи моделі що прогнозувала кількість людей, які відвідають сеанс.

Для цього з історії було взято розклад одного з тижнів прокату фільмів.

Для тестування було завантажено:

- Реальний відсоток наповненості залів (рис. 4.3);
- Прогноз моделі що до наповненості залів на існуючий розклад (рис. 4.4);





Рисунок 4.3 – Реальна кількість відвідувачів кінотеатру на сеанс



Рисунок 4.4 – Кількість відвідувачів кінотеатру на сеанс за прогнозом

Перевірка виявила що моделі стали краще оцінювати заповненість залів а ніж кількість людей, як це було в минулій роботі (табл. 4.1). Це – логічно, так як моделі такого типу набагато краще працюють з нормованими даними, а ніж з звичайними числами.



Таблиця 4.1 – Оцінки моделей першої частини

	MAPE	MSE
LightGBM	0.32	250.6
LSTM	0.25	145.8

Стосовно другої частини комплексу, були отримані дуже цікаві висновки. Аналіз важливості параметрів моделі показав наступні результати (рис. 4.5):

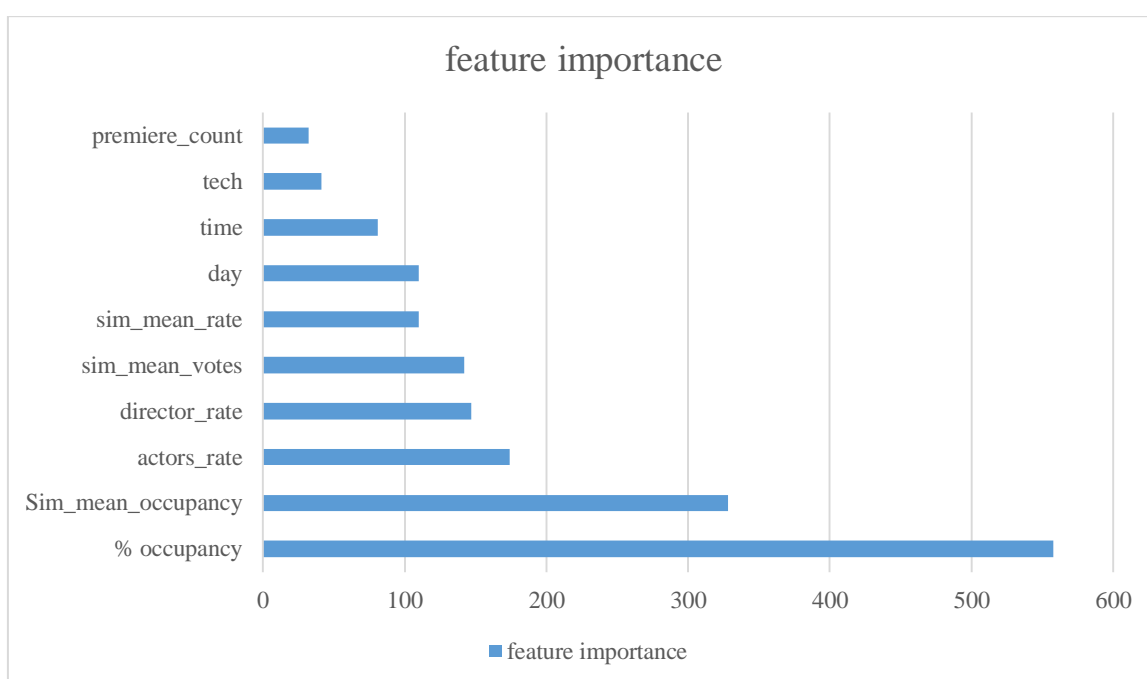


Рисунок 4.5 – Аналіз важливості параметрів моделі

Найбільш важливим показником, як і очікувалось, став відсоток заповненості сеансу, наступний за ним це показник заповненості схожих фільмів, далі рейтинги акторів та режисерів фільму прем'єри оцінки imdb і в самому кінці день показу, час сеансу.

Показник – кількість прем'єр одночасно, неочікувано став найменш важливим у моделі, але зовсім його прибирати не можна, так як від цього якість моделі падає.

Всі інші показники вписуються в очікувані результати так як представник замовника підтвердив що в першу чергу на розміщення сеансів впливають саме схожі фільми та їх показники, а також акторський склад фільму.

Було виявлено що створений програмний продукт ефективно проектує план сеансів на день, враховуючи всі необхідні нам параметри по фільмах (табл. 4.2, табл. 4.3).

Таблиця 4.2 – Оцінки якості моделі

target	Precision	recall	F1-score	support
0*	0,86	0,84	0,85	212
1*	0,85	0,86	0,86	219
accuracy			0,85	

Примітка. \* – 0 це сеанс не потрапив у заплановані, а 1 – потрапив.

Таблиця 4.3 – Показники TP, FN, FP, TN для базового порогу довіри

real	predict	count
0*	0*	179
	1*	33
1*	0*	30
	1*	189

Примітка. \* – 0 це сеанс не потрапив у заплановані, а 1 – потрапив.

Також було проведено експеримент для порогу точності, так як базовий поріг був рівний 0,5 (рис. 4.6).

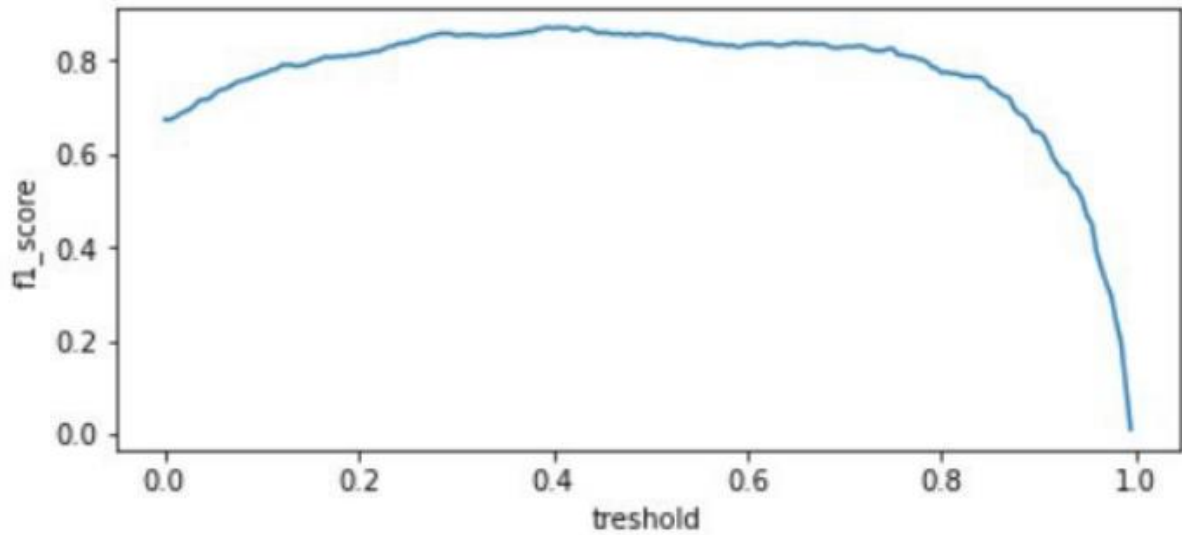


Рисунок 4.6 – Експерименти з порогом довіри

Після отримання результатів, був проведений ще один тест якості (табл. 4.3)

Таблиця 4.4 - Показники TP, FN, FP, TN для експериментального порогу довіри

real	predict	count
0*	0*	167
	1*	45
1*	0*	15
	1*	204

Примітка. \* – 0 це сеанс не потрапив у заплановані, а 1 – потрапив.

## Висновки до розділу 4

Тестування створеного програмного продукту було проведено в три етапи.

За результатами експертної оцінки першого етапу модель підбору схожих фільмів була визнана робочою.

За результатами другого етапу було з'ясовано, що комплекс моделей успішно справляється з задачею прогнозування наповненості залів з достатньою точністю.

За результатами третього етапу було виявлено, що створений програмний продукт ефективно проектує план сеансів на день, враховуючи всі необхідні нам параметри по фільмах.

## РОЗДІЛ 5. РОЗРОБКА СТАРТАП-ПРОЕКТУ

### Вступ

Тенденція останніх років засвідчила, що зростаюча кількість малих інноваційних стартап-компаній, які працюють через Інтернет-платформи, формують майбутнє бізнесу. Тим не менш, статистика засвідчує, що дев'ять з десяти стартапів закінчуються невдачею. Вчені дослідили, що основною причиною невдач (близько 42% випадків відмови) є відсутність попиту на створений товар чи послугу. Ефективним інструментом для запобігання провалу проекту на початкових етапах є ретельне дослідження ринку та складання бізнес-плану, який визначає бізнес-ідею, управлінський підхід та бізнес-стратегію. Для розробки стартап проекту та виведення його на ринок необхідно провести детальне дослідження, яке передбачає виконання таких пунктів:

- 1) здійснити маркетинговий аналіз стартап-проекту:
  - а. розробити опис ідеї проекту, визначити основні напрямки використання товару чи послуги та сформулювати основні відмінності від товарів/послуг конкурентів;
  - б. проаналізувати ринкові можливості для його реалізації;
  - в. розробити стратегію виведення товару на ринок базуючись на аналізі ринкового середовища.
- 2) організація стартап-проекту:
  - а. скласти календарний план реалізації та запуску стартап-проекту;
  - б. визначити плановий обсяг виробництва потенційного товару та на його основі розрахувати потребу у матеріальних ресурсах і персоналі;
  - в. розрахувати витрати, необхідні для реалізації проекту, та витрати на запуск проекту.
  - г. виконати фінансово-економічний аналіз та оцінити ризики стартап-проекту, в межах якого:
  - д. визначити обсяг інвестиційних витрат;

- е. розрахувати основні фінансово—економічні показники проекту (собівартість, ціну продукту/послуги, податковий збір та чистий прибуток) та визначити показники інвестиційної привабливості проекту (рентабельність продажів, період окупності проекту);
  - ж. визначити основні ризики проекту та способи для їх запобігання.
- 3) розробити заходи з комерціалізації проекту:
- а. визначити цільову групу інвесторів та описати їх бізнес інтереси;
  - б. скласти інвестиційну пропозицію;
  - в. стислий опис проекту для ознайомлення інвестора із стартап-проектом;
  - г. визначити основні канали та заходи для просування oferty інвесторам.

## 5.2 Опис ідеї стартапу

Ідея товару генерується відповідно до такого розширеного алгоритму:

- Визначення функції, які повинен виконувати прийнятний варіант виробу.
- Подання на карті широкого спектра елементарних рішень(табл. 5.1), тобто альтернативних засобів реалізації кожної функції.
- Обрання по одному прийнятному елементарному рішенню для кожної функції.

Ідея: Комплекс автоматизованого формування розкладу для кінотеатрів .

Для кого : люди що займаються генерацією розкладів у кінотеатрах.

Основні функції:

- можливість зберігати розклади;
- швидкодія;
- підтримка спеціалістів;

– мультисистемність.

Таблиця 5.1 – Спектр рішень

Спектр рішень	1-ше	2-ге	3-те	4-те
можливість зберігати розклади	Диск комп'ютера	хмарне сховище	Власний аккаунт	інше
Швидкодія	можливість налаштування ресурсоемності	декілька процесорів	інше	інше
підтримка спеціалістів	Залучення спеціаліста розробки на погодинну оплату	Залучення спеціаліста розробки на повну ставку	інше	інше
мультисистемність	Веб-сторінка для браузерів	Браузери+and roid додаток	Браузери+Apple додаток	інше

Обрання по одному прийнятному рішення для кожної функції. Темним затіненням показано традиційну систему схожих продуктів, більш світлим – інноваційний варіант.

Переваги інноваційної розробки:

- використання власного аккаунту для генерації та збереження розкладу;
- автономність кожного окремого користувача;
- можливість повного налаштування системи під кожен окремий кінотеатр;
- зручність використання.

Таким чином, ідею нового товару можна сформулювати так: програмний продукт для автоматичної генерації розкладу сеансів складається з модулів, як

базової комплектації (для генерації розкладу лише у аккаунті) так і для підвантаження в cgm систему кінотеатру.

Орієнтовна вартість 200 000 грн.

Таблиця 5.2 – Товар за задумом

Задум товару	Опис
Товар за задумом	Програмний комплекс вигідний для кінцевого споживача (укладача розкладу) так як значно зменшує час на щоденний аналіз контенту та ручне розподілення сеансів. Товар є конкурентоспроможним.
Товар у реальному виконанні	Програмний комплекс для генерації розкладу розміщено н сторінці браузеру. Для кожного клієнта створюється акаунт де він в подальшому зможе працювати. вартість 200 000 грн
Товар з підкріпленням	Дизайн програмного комплексу може бути оформлений відповідно за побажанням замовника. За його бажанням можна також додати якісь персональні функції. Можливий продаж у кредит. Можлива реалізація товарів для компанії зі знижками.



Таблиця 5.3 – Питання поставлені до проекту

№	Поставлене питання	Відповідь
1	Продукт	Програмний комплекс для автоматичної генерації розкладу
2	Частиною яких систем є продукт?	Систем моніторингу та контролю якості надання послуг у кінотеатрах
3	Чи можна розділити продукт на частини?	Модулі прогнозування та вивантаження розкладу
4	Чи можна об'єднати (агрегувати) кілька з них?	Ні
	Яким має бути ідеальний продукт?	Укладач розкладу не має наміру змінювати щось в згенерованому розкладі
	Що відбудеться, якщо вилучити цей продукт?	Наразі всі маніпуляції з розкладом проводяться вручну
	На розвиток яких функцій було спрямоване удосконалення продукту	Головним чином розвиток технологій було спрямовано на зменшення часу роботи людини в цій сфері.
	Які функції залишилися «недорозвиненими»?	Відсутність впливу реакцій людей на фільм в прокаті
	Які функції залишилися «недорозвиненими»?	Встановити системи моніторингу облич на виході з залу щоб отримувати оцінку фільму глядачами

### 5.3 Проектування стартапу та формування команди

Визначаємо основні елементи вкладу в створення стартапу. Demmler пропонує наступні:

- ідея;
- підготовка бізнес-плану;

- компетентність;
- залученість і ризики;
- обов'язки.

Визначення важливості елементів щодо їх вкладу у створення та реалізацію стартапу (табл. 5.4).

Таблиця 5.4 – Визначення важливості елементів стартапу

Параметр	Вага (важливість)
Ідея	7
БП	9
Компетентність	6
Залученість	6
Обов'язки	7
Залучення партнерів	9

Оцінювання особистого внеску кожного партнера у створення та реалізацію стартапу (табл 5.5).

Таблиця 5.5 – Визначення особистого внеску кожного партнера

Елементи	PM\BI	Dev	Dev	Test	DS
Ідея	5	3	2	2	8
БП	9	0	0	0	4
Компетентність	4	6	8	6	8
Залученість	5	9	8	6	9
Обов'язки	8	6	6	6	8
Залучення партнерів	10	0	0	0	2

Оцінювання важливості кожного фактора і внеску кожного учасника (табл. 5.6).

Таблиця 5.6 – Оцінювання важливості кожного фактора і внеску кожного учасника

	PM\BI	Dev	Dev	Test	DS	Варг
<b>Ідея</b>	5	3	2	2	8	7
<b>БП</b>	9	0	0	0	4	9
<b>Компетентність</b>	4	6	8	6	8	6
<b>Залученість</b>	5	9	8	6	9	6
<b>Обов'язки</b>	8	6	6	6	8	7
<b>Залучення партнерів</b>	10	0	0	0	6	9

Визначення дольової участі у стартап проєкті кожного учасника (табл. 5.7).

Таблиця 5.7 – Визначення дольової участі кожного учасника

	PM\BI	Dev	Dev	Test	DS	Всього
<b>Ідея</b>	5	3	2	2	8	17
<b>БП</b>	9	0	0	0	4	13
<b>Компетентність</b>	4	6	8	6	8	32
<b>Залученість</b>	5	9	8	6	9	37
<b>Обов'язки</b>	8	6	6	6	8	34
<b>Залучення партнерів</b>	10	0	0	0	2	12
<b>Всього</b>	41	24	24	20	39	145
	<b>28,07%</b>	<b>16,03%</b>	<b>16,24%</b>	<b>13,33%</b>	<b>26,33%</b>	<b>100%</b>

Висновок: Якби система була досить простою, то будь-який розробник міг би її побудувати в лічені дні, проте, дійсно потрібен широко профільний спеціаліст та бізнес-аналітик, котрі зможуть структурувати та розробити бізнес-план створення такого програмного продукту. При всій простоті калькулятор Деммлера допомагає більш-менш об'єктивно оцінити внесок кожного учасника стартапа і належну йому частку без зайвих суперечок. При цьому, в даній ситуації калькулятор досить вірно розраховує початковий внесок усіх партнерів у проект.

## 5.4 Розроблення MVP

Головна проблема: потрібна якісна та ефективна система для автоматичної генерації розкладу сеансів кінотеатрів.

### MVP1

Ручна обробка даних, кластеризація фільмів вручну, побудова планів сеансів, формування рекомендацій для реклами

### MVP 2

Автоматизована обробка даних вбудованими скриптами powerbi, менше часу на аналіз фільмів, ручна генерація розкладів

### MVP3

Проста система на конструкціях if-else, яка дозволяє автоматично будувати аналіз фільмів в прокаті, виводити оцінку популярності фільмів, ручна генерація розкладу

### MVP4

Система прогнозування кількості людей на кожен окремий сеанс, збудована локально, залита на сервер, запити до неї через консоль, відбудова та вивантаження через streamlit, напівавтоматична генерація розкладу

### MVP5

Веб–сервіс з можливістю перебудовувати моделі та використовувати різні джерела даних, з можливістю вивантаження результатів в багатьох форматах, використання супутніх даних про фільми для прогнозу, автоматична генерація розкладу

### 5.5 Аналіз ринкових можливостей та розроблення маркетингової стратегії стартап-проекту

Характеристика потенційних клієнтів стартап-проекту представлена в таблиці 5.8.

Таблиця 5.8 – Характеристика потенційних клієнтів стартап-проекту

№	Потреба що формує ринок	Цільова аудиторія (цільові сегменти ринку)	Відмінності у поведінці різних потенційних цільових груп клієнтів	Вимоги споживачів до товару
1	Аналіз фільмів на прибутковість	Власники кінотеатрів	Зацікавлені в формуванні прибуткових розкладів	Простий інтерфейс, гнучкі налаштування
2	Прогноз кількості людей на сеанси	Власники кінотеатрів	Зацікавлені в забезпеченні високої виручки за рахунок знань про можливу кількість людей на кожен сеанс	Можливість виділяти сегменти швидкого реагування на зміни в моделях, сервер для збереження даних, візуалізація, розсилка реклами
3	Автоматична генерація розкладу	Укладачі розкладу, власники кінотеатрів	Зацікавлені в автоматичній генерації розкладу з можливістю його коригування за потреби	Можливість виділяти сегменти швидкого реагування, візуалізація, розсилка реклами, Коригування розкладу, вивантаження в CRM системи

Визначення ринкових можливостей і загроз представлено в таблиці 5.9.

Таблиця 5.9 – Визначення ринкових можливостей і загроз

Параметри оцінки	Можливості	Загрози
Конкуренція	Відсутнє повноцінне ПО на ринку	Ідеї такого продукту вже давно є у розробках інших компаній
Збут	Можливість налаштування під конкретного клієнта	Багато систем ведуть свої бази по різному

Ступеневий аналіз конкуренції на ринку представлено в таблиці 5.10.

Таблиця 5.10 – Визначення ринкових можливостей і загроз

Особливості конкурентного середовища	В чому проявляється дана характеристика	Вплив на діяльність підприємства (можливі дії компанії, щоб бути конкурентоспроможною)
Вказати тип конкуренції – монополія	Відсутні конкуренти (поки що)	Виконати мінімальний функціонал якомога швидше
За рівнем конкурентної боротьби – локальний	Розробка ведеться виключно для України	
За галузевою ознакою внутрішньогалузева	Підходить до конкретної галузі	Розробити базову систему для декількох галузей
За характером конкурентних переваг – Інноваційна	Використання нових технологій в аналізі	

Аналіз конкуренції в галузі за М. Портером представлено в таблиці 5.11.

Таблиця 5.11 –Аналіз конкуренції в галузі за М. Портером

Складові аналізу	Прямі конкуренти в галузі	Потенційні конкуренти	Постачальники	Клієнти	Товари-замінники
	Навести перелік прямих конкурентів	Визначити бар'єри входження в ринок	Визначити фактори сили постачальників	Визначити фактори сили споживачів	Фактори загроз з боку замінників
Висновки :	відсутні	Невідомо на яких етапах ведуться розробки в компанії	Проект залежить від даних постачальників	Проект залежить від даних постачальників	відсутні

Обґрунтування факторів конкурентоспроможності представлено в таблиці 5.12.

Таблиця 5.12 – Обґрунтування факторів конкурентоспроможності

Фактор конкурентоспроможності	Обґрунтування (наведення чинників, що роблять фактор для порівняння конкурентних проектів значущим)
Інноваційні рішення	Використання технологій DS робить фінальний продукт бажаним для кінцевого споживача
Широкий спектр послуг	Проект є модульним, з можливою добудовою функціоналу під побажання клієнта

SWOT– аналіз стартап-проекту представлено в таблиці 5.13.

Таблиця 5.13 – SWOT- аналіз

<b>Сильні сторони:</b> Рішення в DS для виходу на ринок Сильна команда	<b>Слабкі сторони:</b> Залежність від даних клієнта Клієнт не завжди розуміє що хоче
<b>Можливості:</b> Відсутність конкурентів	<b>Загрози:</b> Розвиток технологій у можливих конкурентів

### 5.6 Формування маркетингової програми та плану маркетингу стартапу

В наступних таблицях (табл. 5.14, 5.15) викладено інформацію про формування маркетингової програми та план маркетингу стартапу.

Таблиця 5.14 – Визначення базової стратегії розвитку

<b>Стратегія охоплення ринку</b>	<b>Ключові конкурентоспроможні позиції відповідно до обраної альтернативи</b>	<b>Базова стратегія розвитку*</b>
Ексклюзивний розподіл	Позиціонування за співвідношенням "ціна – якість", позиціонування на сервісному обслуговуванні, позиціонування на позитивних особливостях технології	Стратегія диференціації



Таблиця 5.15 – Визначення базової стратегії конкурентної поведінки

Чи є проект «першопрохідцем» на ринку?	Чи буде компанія шукати нових споживачів, або забирати існуючих у конкурентів?	Чи буде компанія копіювати основні характеристики товару конкурента, і які?	Стратегія конкурентної поведінки*
Так, аналоги продукту відсутні	Так	Ні	Стратегія Лідера / Креативна.

Креативна стратегія – компанія безпосередньо буде просувати безліч інновацій, а також забезпечувати повну орієнтацію на споживача, а саме підтримувати повне обслуговування, матиме гарантію функціонування та забезпечить різноманітні акції для користувачів. Особливо вирізнятиме компанію постійне оновлення ПО, та впровадження покращень для користувачів, компанія буде постійно вдосконалюватись та орієнтуватись на користувачів, так, наприклад, знайдені помилки в ПО будуть в першу чергу виправлятись, а відповідний користувач, що знайшов матиме певні переваги, така стратегія допоможе покращити продукт та запровадить ініціативу та зацікавленість користувача.

Ключові переваги концепції потенційного товару та опис трьох рівнів моделі товару представлено в наступних таблицях (табл. 5.16 , 5.17).

Таблиця 5.16 – Визначення ключових переваг концепції потенційного товару

Потреба	Вигода, яку пропонує товар	Ключові переваги перед конкурентами (існуючі або такі, що потрібно створити)
Забезпечення розсилки реклами, забезпечення аналізу продажів, забезпечення прогнозування кількості проданих квитків, забезпечення моніторингу наповненості залів	Продукт дає змогу компанії моніторити та прогнозувати необхідні ресурси компанії	Постійна підтримка ПО Можливість доробки ПО під клієнта

Таблиця 5.17 – Опис трьох рівнів моделі товару

Рівні товару	Сутність та складові
I. Товар за задумом	Керування процесами генерації розкладу, моніторинг ресурсів компанії.
II. Товар у реальному виконанні	Властивості/характеристики <ol style="list-style-type: none"> <li>1. Автоматизована система керування</li> <li>2. Інтеграція з усіма ОС</li> <li>3. Автоматичні оновлення</li> </ol>
III. Товар із підкріпленням	До продажу: дослідження ефективності на тестових кейсах, тестування, постійні оновлення
	Після продажу: наявність постійних апдейтів програмного забезпечення
За рахунок чого потенційний товар буде захищено від копіювання:	Ліцензія сертифікація наявність trial version перед купівлею.

Визначення меж встановлення ціни, формування системи збуту та концепцію маркетингових комунікацій представлено в наступних таблицях (табл. 5.18, 5.19, 5.20).

Таблиця 5.18 - Визначення меж встановлення ціни

Рівень цін на товари-замінники	Рівень цін на товари-аналоги	Рівень доходів цільової групи споживачів	Верхня та нижня межі встановлення ціни на товар/послугу
5000 \$	3000-5000 \$	90000 \$ / рік	5000-6000 \$ / рік

Таблиця 5.19 - Формування системи збуту

Специфіка закупівельної поведінки цільових клієнтів	Функції збуту, які має виконувати постачальник товару	Глибина каналу збуту	Оптимальна система збуту
Клієнти купують продукт. Також можливе постійне придбання апдейтів та спеціального функціоналу.	Забезпечити безпечну грошову транзакцію, непошкодженість та робочий стан всіх елементів системи, текстову та відео інструкцію.	Трирівневий канал збуту	Зовнішні засоби збуту

Таблиця 5.20 - Концепція маркетингових комунікацій

Специфіка поведінки цільових клієнтів	Канали комунікацій, якими користуються цільові клієнти	Ключові позиції, обрані для позиціонування	Завдання рекламного повідомлення	Концепція рекламного звернення
Цільові клієнти не мають та не користуються подібними пристроями, так як він є унікальним в своєму роді, однак можливо користуються окремими компонентами системи.	Рекомендації знайомих, прямий зв'язок з цільовими клієнтами, соціальні зв'язки в мережі Інтернет	Позиціонування на основі порівняння товару фірми з товарами конкурентів, Позиціонування на позитивних особливостях технології, Позиціонування за показниками якості, Позиціонування як унікального продукту та функціоналу.	Привернути увагу споживача, ознайомити з усіма перевагами використання продукту, переконати в унікальності продукту, показати усі переваги використання продукції, показати усі переваги клієнта компанії.	Презентація переваг та унікальних характеристик пристрою. Наголос на чудових результатах досліджень та тестування. Пропозиція тестування продукції. Наголос на перевагах клієнта, акціях та повній підтримці компанії.

## Висновки до розділу 5

У результаті виконання даного розділу було з'ясовано, що існує реальна можливість ринкової комерціалізації розробленого продукту.

Даний продукт буде рентабельним, оскільки на україномовному ринку відсутні аналоги і конкуренція.

Сервіс матиме стандартний набір функціоналу, що задовольнятиме всі потреби клієнтів, а за необхідності його можна буде доповнювати.

## ВИСНОВКИ

У роботі було проаналізовано створену раніше систему формування розкладу сеансів для кінотеатрів, виявлено слабкі місця системи розподілу сеансів, особливо що стосувалось прогнозу фільмів з історією продажів.

Було розроблено архітектуру моделей для автоматизованого формування розкладу сеансів для кінотеатрів, що включає в себе окремі моделі для прогнозування заповненості залу на сеанс в залежності від кількості часу прокату фільму.

Створено програмний комплекс автоматизованого формування розкладу сеансів для кінотеатрів, враховуючи дані фільмів.

Запропоновано метод для автоматичного формування розкладу сеансів кінотеатрів, який відрізняється від існуючих тим що не потребує втручання людини на етапі прогнозування популярності фільму в прокаті та дозволяє лише коригувати отриманий розклад за потреби, не створюючи його з нуля.

Протестовано програмний комплекс автоматизованого формування розкладу сеансів для кінотеатрів на реальних даних в оточенні замовника.

При проведенні маркетингового аналізу було з'ясовано, що існує реальна можливість ринкової комерціалізації розробленого продукту.

Даний продукт буде рентабельним, оскільки на україномовному ринку є лише часткові аналоги та відсутня конкуренція.

## ПЕРЕЛІК ПОСИЛАНЬ

1. Обзор методов прогнозирования. URL: <https://ivan-shamaev.ru/overview-forecast-methods/> (дата звернення: 03.05.2020).
2. Mathematical model. URL: [https://www.sciencedaily.com/terms/mathematical\\_model.htm](https://www.sciencedaily.com/terms/mathematical_model.htm) (дата звернення 03.05.2020)
3. Чучуева И. А. Модель прогнозирования временных рядов по выборке максимального подобия: дис. канд. тех. наук 05.13.18/ Московский государственный технический университет им. Н. Э. Баумана. Москва, 2012. 155 с.
4. Тихонов Э. Е. Прогнозирование в условиях рынка. Невинномысск: ЗАО «Невинномысская городская типография», 2006. 221 с.
5. Мерков А. Б. Распознавание образов: Введение в методы статистического обучения. Монография. Москва: Едиториал УРСС, 2011. 254 с.
6. Сидоров С. Г., Никологорская А. В. Анализ временных рядов как метод построения потребления электроэнергии / *Вестник ИГЭУ*. 2010, Вып. 3. С. 81–83.
7. Armstrong J.S. Forecasting for Marketing. Quantitative Methods in Marketing. London: International Thompson Business Press, 1999. 119 p.
8. Бокс Дж., Дженкинс Г. М. Анализ временных рядов, прогноз и управление. Москва: Мир, 1974. 406 с.
9. Фейгин Л. И. Задачи теории расписаний при нечетких длительностях операций. Доклады АН СССР. 1983. Т. 272. №4. С. 812–815.
10. Шеннон Р. Имитационное моделирование систем. Искусство и наука: пер. с англ. Москва: Мир, 1978. 418 с.
11. Хаббард Дж. Автоматизированное проектирование баз данных. Москва: Мир, 1983. 295 с.
12. Щербакова И. В. Математическое моделирование информационных систем центров ситуационного управления в интересах обеспечения

- безопасности: автореф. дис.канд. техн. наук. Воронеж: ВИ МВД России, 2009. 16 с.
13. Гудфеллоу Я., Бенджио И., Курвилль А. Г93 Глубокое обучение / пер. с англ. А. А. Слинкина. – 2-е изд., испр. Москва: ДМК Пресс, 2018. 652 с.
  14. Матвиенко Ю.А., Зинченко А.А. Разработка программного обеспечения для оптимизации методики алергодиагностики к стоматологическим анестетикам материалы 20-й Международной научно-технической конференции SAIT 2018, Киев, 21 – 24 мая 2018 г. УНК “ИПСА” НТУУ “КПИ им. Игоря Сикорского”. Киев: УНК “ИПСА” НТУУ “КПИ”, 2018. С. 247-248.
  15. Матвієнко Ю. О., Рекалова О. М., Тлустова Т. В., Зінченко А. О. Використання методу седиментації еритроцитів для діагностики непереносимості протитуберкульозних препаратів у хворих на туберкульоз легень/ *Астма та алергія*. 2020. № 2. С. 57–66.
  16. Технология определения медикаментозной непереносимости по иммунологическому тесту седиментации эритроцитов с противотуберкулезными препаратами у больных туберкулезом легких / Ю. А. Матвиенко и др. // Материалы междунар. научно-практической конференции «Современные вопросы диагностики, лечения и профилактики туберкулеза». Баку: 2018. С. 92.

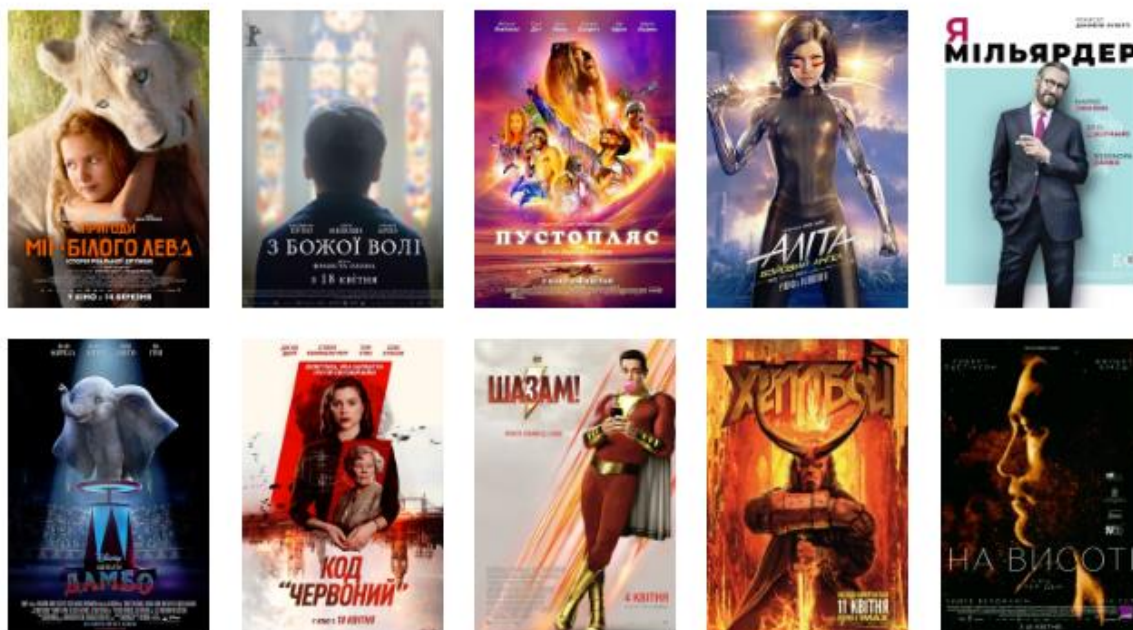


## ДОДАТОК А.

## Обери фільм

Пошук

## Зараз в прокаті



## Випадково обрані фільми



Рисунок А.1 – Поле вибору фільму для аналізу

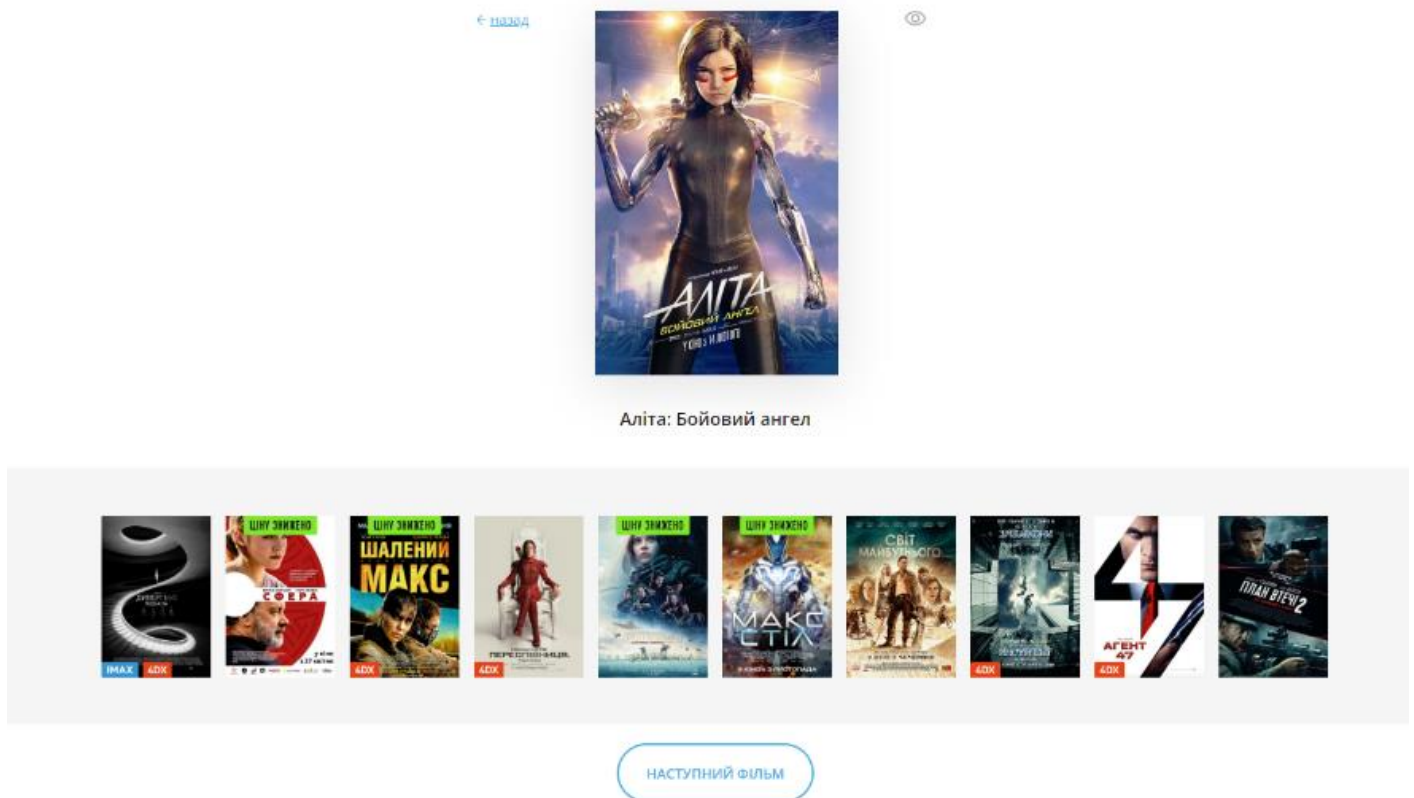


Рисунок А.2 – Вікно аналізу фільмів

Создание

Код IMDb

tt0437086

ЗАПОЛНИТЬ

Название

Аліта: Бойовий ангел

Оригинальное название

Alita: Battle Angel

Продолжительность

122

Цвет

Возрастное ограничение

PG-13

Режиссёр

Robert Rodriguez

Жанры

Action

Adventure

Animation

Biography

Актёры

Rosa Salazar, Christoph Waltz, Jennifer Connelly, Mahershala Ali

Для множественного выбора удерживайте клавишу CTRL

ОТМЕНА

СОХРАНИТЬ

Рисунок А.3 – Вікно вводу обмежень дистриб'юторів

Редактирование

Фильм

Аліта: Бойовий ангел

Кинотеатры

Кинотеатр 2

Кинотеатр 3

Кинотеатр 4

Кинотеатр 1



Для множественного выбора удерживайте клавишу CTRL

Показы за день

Показы за период

Минимальные цены

ДОБАВИТЬ

Начало	Окончание	Технология	Количество	
2019-04-01	2019-04-30	IMAX3D	2	 

ОТМЕНА

СОХРАНИТЬ

Рисунок А.4 – вікно збору даних по фільму